الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Dr. YAHIA FARES de Médéa

Faculté des Sciences
Département Mathématiques et Informatique



Mémoire de fin d'études de Master

Présenté pour l'obtention du Diplôme de Master en Informatique

Option : Ingénierie des systèmes et technologie du web

THEME

Discrimination d'une voix pathologique par la classification Deep-Learning basée images. Etude de Cas : Maladie de Parkinson

Proposé et dirigé par :

Melle. Djamila MEGHRAOUI

Mr. Ahmed Cherif MAZARI

Préparé et présenté par :

Melle. Nihal BELLIL

Année Universitaire 2019-2020

Remerciements

- Je souhaite manifester mes sincères remerciements à Dieu, le tout puissant pour ses faveurs et ses gratitudes de m'avoir donné le courage et la patience afin d'achever ce modeste travail.

- Je tiens à présenter de tout mon cœur mes remerciements et reconnaissances à mes honorables encadreurs M^{elle} «MEGHRAOUI Djamila» et M^{r} «MAZARI Ahmed Cherif» pour :

  - leur aide,
  - leurs conseils précieux,
  - leur gentillesse,
  - leur encouragement, sa disponibilité et leur confiance qui m'a permis de ne jamais faiblir et de poursuivre toujours plus loin dans mon travail.

- Je remercie naturels à tous mes professeurs de l'école primaire jusqu'à l'université pour leur soutien moral, leurs conseils et leur éducation dans tous nos parcours académiques, ainsi que leur générosité et leurs encouragements qui ont été pour moi une source de courage et de confiance.

- Je remercie aussi vivement les honorables membres du jury qui ont accepté d'évaluer ce travail.

- Enfin, un grand remerciement à tous ceux qui, par un mot, m'ont donné la force de continuer à travailler afin d'atteindre mes objectifs.

Dédicaces

Je dédie ce mémoire

À mes chers parents ma mère et mon père

À toute ma famille

À mes amis

À tous ceux qui, par un mot, m'ont donné la force

De continuer…

Résumé

L'efficacité des systèmes de reconnaissance vocale est importante pour une bonne interaction homme-machine. En raison du manque de données audio au moment du développement massif de l'apprentissage profond, nous avons proposé une méthode basée sur l'augmentation des données en utilisant les coefficients cestreaux fréquence Mel (MFCCs). Ces coefficients sont convertis par la suite en images afin de les classifier par un model d'apprentissage profond. Ensuite, pour tester la performance de l'approche proposée, on fait appel aux réseaux de neurones convolutifs conventionnels (CNN), développés par VGG16, INCEPTIONV3 et RESNET. Les résultats obtenus ont montré l'efficacité de l'approche proposée pour aider le diagnostic d'une maladie neurologique (maladie de Parkinson comme étude de cas) en utilisant la voix.

Mots clés : Apprentissage profonds, MFCC, maladie neurologique, la voix

Abstract

The efficiency of speech recognition systems is important for good human-machine interaction. Due to the lack of audio data at the time of the massive development of deep learning, we proposed a method based on data augmentation using Mel frequency cepstrum coefficients (MFCCs). These coefficients are later converted into images in order to classify them by a deep learning model. Then, to test the performance of the proposed approach, the conventional convolutive neural networks (CNN) developed by VGG16, INCEPTIONV3 and RESNET is used. The results demonstrated the effectiveness of the proposed approach to assist in the diagnosis of a neurological disease (Parkinson's disease as a case study) using voice.

Keys: deep learning, MFCC, neurological disease, the voice.

الملخص

تعد كفاءة أنظمة التعرف على الصوت مهمة للتفاعل الجيد بين الإنسان والآلة. و نظرا لنقص البيانات الصوتية في وقت التطور الهائل للتّعلم العميق. اقترحنا في هذا العمل طريقة تعتمد على زيادة البيانات باستخدام معاملات MFCCs و تحويلها إلى قاعدة صور . هذه الصور تم إدخالها كمعطيات للنموذج المقترح بغرض معالجتها وتصنيفها باستعمال التعلّم العميق. من اجل تقييم الفكرة المقترحة تم استعمال الشبكات العصبية التلافيفية التقليدية CNNs و المطورة VGG16 و INCEPTIONV3 و RESNET .النتائج المتحصل عليها بيّنت إمكانية الخطة المقترحة في هذا العمل على المساعدة في تشخيص مرض الباركنسون كدراسة حالة باستعمال الصوت.

المفتاح : التعلم العميق ، MFCC ، الأمراض العصبية ، الصوت.

Table des matières

Introduction Générale..... 11

CHAPITRE I : La Production De La Parole

1. Introduction 19
2. Anatomie de la parole 19
   2.1 Voix humaine 19
   2.2 Production de la voix 19
   2.3 Appareil phonatoire humain 20
   2.3.1 Appareil respiratoire 21
   2.3.2 Larynx 22
   2.3.2.1 Définition (généralité) 22
   2.3.2.2 Structure des cordes vocales 23
   2.4 Cavité résonante 24
   2.4.1 Le pharynx 25
   2.4.2 Cavité buccale 25
   2.4.3 Cavité labiale 26
   2.4.4 Fosse nasale (Cavité nasale) 26
   2.4.5 Cavité orale 27
3. Pathologie de la parole : Dysarthrie Parkinsonienne 27
   3.1 Introduction sur la maladie de Parkinson 28
   3.2 Historique 30
   3.3 Dysarthrie Parkinsonienne 31
   3.3.1 Causes de la maladie Parkinson 31
   3.3.2 Symptômes de parkinson 32
   3.3.3 Diagnostic de Parkinson 33
4. Les méthodes d'évaluation d'une dysarthrie 33
   4.1 Preceptive 33
   4.2 Objective 34
5. Conclusion 35

CHAPITRE II : Analyse Et Discrimination Automatique

1.Introduction 41
2.Analyse acoustique et discrimination automatique 41
2.1 Introduction (généralité) 41
2.2 Paramètres MFCCs 42
2.2.1 Définition 42
2.2.2 Étapes de l'extraction des paramètres MFCCs 42
2.3 Discrimination automatique 46
2.3.1 Deep Learning 46
2.3.1.1 Définition(généralité) 46
2.3.1.2 Avènement du Deep Learning 46
2.3.1.3 Origine de Deep Learning 46
2.3.2 De Machine Learning au Deep Learning 46
2.3.3 Fonctionnement de Deep Learning 47
2.3.4 Exemple de Deep Learning 48
2.3.4.1 Conduite automatisé 48
2.3.4.2 Défense et aérospatial Deep Learning 48
2.3.4.3 Recherche médicale 48
2.3.4.4 Automatisation industrielle 49
2.3.4.5 Traduction 49
2.3.4.6 Reconnaissance vocale au niveau quasi humain 49
2.3.5 Les Algorithmes de Deep Learning 49
2.3.5.1 Réseau de neurones 50
2.3.5.2 Réseau Neuronal Convolutif (CNN) 50
2.3.6 Réseaux de neurones convolutifs pour la classification d'images 54
2.3.7 Apprentissage Par Transfert 55
2.3.7.1 Modèle VGG16 56
2.3.7.2 Modèle Xception 56
2.3.7.3 Modèle Inception V3 57
2.3.7.4 Modèle ResNet 57
3. Mesures d'évaluation des modèles 58
4. Classification Des Images ..... **Erreur ! Signet non défini.**

5. Travaux récents utilisant Deep Learning pour la parole 61
6.Conclusion 62

CHAPITRE III : Approche De Travail

1.Introduction 73
2.Approche proposée 73
2.1. Schéma du Modèle proposé 73
3.Base de données 75
3.1. Définition (généralité) 75
4.Extraction des paramètres MFCCs 75
5.Classification 76
5.1 Définition (généralité) 76
5.2 Méthode proposée pour la classification des images 76
5.3 Architecture du Modèle CNN proposé 77
5.4 Architecture du Modèle Apprentissage par transfert proposé 79
6.Conclusion 82

CHAPITRE IV : Test Et Résultats

1.Introduction 87
2.Les outils et les librairies utilisés 87
2.1. TensorFlow 87
2.2 Keras 87
2.3. Python 88
2.4 Configuration utilisé dans l'implémentation 88
3. Description de la Base de données 89
3.1 Base de données originale 89
3.2 La modélisation des fichiers sons en images 90
4.Expérimentations et résultats 93
4.1 Résultat obtenue pour modèle CNN 93
4.2 Paramètre d'évaluation du réseau CNN 94
4.3 Résultat obtenue de Modèle apprentissage par transfert 94
4.4 L'évaluation 96
5.Conclusion 98
Conclusion Générale 95
Références Bibliographiques 97

Liste des Figures

Figure I.1 : Schéma Des organes principaux impliqués dans la production de la parole.....20
Figure I.2 : schéma de l'appareil phonatoire 21
Figure I.3 : Explication de l'appareil respiratoire humain.....21
Figure I.4 : Coupe sagittale de la bouche et du cou, comprenant le larynx et l'appareil supra-laryngé 22
Figure I.5 : schémas de coupe du larynx. 22
Figure .I.6: structure de larynx.....22
Figure I.7 : les positions des cordes vocales. 23
Figure I.8 : cavité buccale.....26
Figure I.9 : la cavité nasale 26
Figure I.10 : cavité oral.....27
Figure I.11 : Maladie de Parkinson 29
Figure I.12 : diminution de la libération de dopamine dans le cas MP 32
Figure II.1 : Étapes d'extraction des coefficients MFCC.....42
Figure II.2 : Hamming Window.....44
Figure II.3 : Les filtres triangulaires passe-bande en Mel-Fréquence.....44
Figure II.4 : Réseau de neurone ET Réseau Deep Learning.....47
Figure II.5 : Classification des méthodes de L'apprentissage profond.....50
Figure II.6 : Schéma d'un réseau de neurones convolutionnel 51
Figure II.7 : Matrices de l'image input de pixels et filtre 52
Figure II.8 : la fonction convolutée.....52
Figure II.9 : la fonction ReLU 53
Figure II.10 : Exemplaire de Max Pooling 53
Figure II.11 : L'architecture de CNN.....54
Figure II.12 : L'architecteur de modèle VGG16 56
Figure II.13 : L'architecteur de modèle Xception 56
Figure II.14 : L'architecteur de modèle Inception V3 57
Figure II.15 : L'architecteur de modèle ResNet.....58
Figure II.16 : matrice de confusion.....59
Figure II.17 : Interprétation graphique de la mesure ROC-AUC. 60
Figure III.1 : Schéma du Modèle proposé.....68
Figure III.2 : modèle VGG16 73
Figure III.3 : Modèle InceptionV3.....75
Figure III.4 : Modèle ResNet.....76
Figure IV.1 : Logo de TensorFlow.....81
Figure IV.2 : Logo de Keras 82
Figure IV.3 : Logo de Python 82
Figure IV.4 : Graphique de mot "ketap" d'un patient non malade 83
Figure IV.5 : Graphiques 20_1814, 20_1829, 20_1834 homme non malade.....84
Figure IV.6 : Graphique de mot "ketap" d'un patient malade.....85
Figure IV.7 : graphique 0020_14174, 0020_15101, 0020_16123 pour femme malade 85
Figure IV.8 : Apprentissage et Validation modèle CNN.....85

Figure IV.9 : matrice de confusion de notre CNN 87
Figure IV.10 : Apprentissage et Validation modèle VGG16 88
Figure IV.11 : Apprentissage et Validation modèle ResNet 89
Figure IV.12 : Apprentissage et Validation modèle InceptionV3 89
Figure IV.13 : Matrice de confusion InceptinV3 92
Figure IV.14 : Matrice de confusion ResNet 92
Figure IV.15: Matrice de confusion VGG16 93

Liste des Tableaux

Tableau III.1: L'architecteur de notre réseau CNN 72
Tableau IV.1: les bibliothèques python utilisé 83
Tableau IV.2: Accuracy des Modèles 91
Tableau IV.3: Rapport de classification des modèles 91

Introduction

Générale

Introduction Générale

La voix est un moyen indispensable par lequel on peut communiquer avec les autres, vu sa permission de transmettre de nombreuses informations, significations et idées différentes. La production d'une voix est un processus d'intégration qui résulte de la coopération des organes vocaux les uns avec les autres. Les organes responsables de la production du son sont appelés : système vocal qui, lui, comprend plusieurs organes dont le système respiratoire, le larynx, les cordes vocales, la cavité supra glottique, le pharynx, la cavité buccale, la cavité buccale et la cavité nasale. Et tout trouble dans ces organes entraîne une perturbation de la voix et une dysarthrie.

La cause des maladies de la voix peut être une infection tissulaire, des changements systémiques, un stress mécanique, une irritation de la surface, des changements tissulaires, des changements neuromusculaires et d'autres facteurs qui les perturbent la fonction et la forme des cordes vocales ce qui entraîne des vibrations irrégulières et une hausse du bruit acoustique. Un tel son semblait être tendu, raide, faible et haletant. Cette anomalie contribue grandement à une mauvaise qualité sonore globale (Katzenshlager et al., 2004). Ces modifications de la voix, prenant la forme de dysarthrie, sont un des premiers symptômes à apparaître dans la maladie de Parkinson (MP).

La maladie de Parkinson touche chaque année près d'un million de personnes dans le monde et représente la deuxième plus grande maladie dégénérative chez les personnes âgées. Dans certains cas, cela peut arriver très tôt, parfois même chez les jeunes de 25 à 30 ans. Les symptômes de la maladie, allant depuis de simples symptômes à la gravité, pouvant devenir très débilitants chez les patients qui en souffrent. Selon le témoignage de l'arsenal thérapeutique à la fois médical et paramédical, nécessaire pour gérer une telle condition.(Rhonda et al., 2000)

La maladie de Parkinson (MP) est une maladie neurologique qui affecte souvent les cellules nerveuses productrices de dopamine dans une zone spécifique du cerveau, provoquant un ensemble complexe de symptômes. Elle est principalement liée à la perte progressive du contrôle moteur du corps.(Katzenshlager et al., 2004) Les principaux symptômes de MP sont:

- tremblements des membres,
- lenteur des mouvements,
- muscles rigides, instabilité de posture,
- perte de mouvements automatiques,
- changement de langage (parler calmement ou rapidement ou confus ou hésiter avant de parler). La maladie de Parkinson est le deuxième trouble dégénératif le plus courant du système nerveux central après la maladie d'Alzheimer (Kibleur & George, 2016).

La plupart du temps, les gens ignorent les premiers symptômes de la MP en tant qu'effet du vieillissement naturel surtout que malheureusement, il n'y a pas de tests médicaux pour détecter la maladie de manière précoce et préventive.

La maladie de Parkinson est donc difficile à diagnostiquer avec précision, cela nécessite également de nombreux frais médicaux qui sont très coûteux car ils nécessitent une prise en charge et une surveillance médicale constantes. Et comme il n'y a pas de remède définitif Il n'est possible que le moyen de réduire la maladie par prescription de certains médicaments. Par conséquent, le patient doit se déplacer chez les spécialistes centres cliniques qui font actuellement défaut à l'Algérie, pour se faire soigner mais les dépenses financières seront importantes compte- tenu du traitement et des déplacements d'un hôpital à un autre ou d'un pays à un autre qui sont très chers. Étant donné que ces moyens de diagnostics et de transports ne sont pas tout le temps disponibles de la difficulté de liquidités financières des non assurés sociaux, il y a ainsi une perte de temps ce qui peut entraîner une exacerbation de la maladie, en particulier les candidats aux derniers stades de la maladie. Pour ces raisons et ces difficultés, une surveillance à distance régulière à des intervalles plus fréquents et à moindre coût a été établie pour suivre la progression de la maladie de Parkinson avec précision et efficacité. Par conséquent, Avec le progrès de la science et la technologie du traitement du qu'une solution efficace et prometteuse pour faciliter le diagnostic par surveillance à distance. (Cheang & Pell, 2007)

Le traitement numérique des signaux sonores est un outil existant qui aide à l'analyse non invasive des cliniciens. Il permet l'identification des troubles vocaux, notamment dès le début.(Muhammad et al., 2017) et la qualité de la voix peut être évaluée par diagnostique ou par un test de laryngoscopie. Deux approches différentes sont incluses: l'approche perceptive et l'approche objective.(Bankar & Gavai, 2018) .

La détection automatique des maladies de la voix peut également être obtenue grâce à différents types d'analyse du signal, qui peuvent être à long ou à court terme. Ces paramètres peuvent être déterminés en utilisant des coefficients cepstraux avec une fréquence Mel(MFCCs).(F. Zheng et al., 2001)

Il existe des classificateurs très importants qui ont été utilisés dans la recherche ancienne pour diagnostiquer les sons pathologiques tels que: -le modèle de Markov caché (HMM) (Deng & Yu, 2013)den plus des réseaux de neurones (Ritchings et al., 2002), -les machines vectorielles de support (SVM) (Deng, 2014) - et enfin le modèle de mélange gaussien (GMM)) (Klabunde, 2002).

Une classification binaire / pathologique normale des échantillons acoustiques (Godino-Llorente et al., 2009) a également été proposée dans la littérature, entre autres.

Dans le discours, la disponibilité des grandes quantités de l'ensemble de données dont le deep learning nécessite beaucoup n'est pas toujours disponible, et si elle est trouvée, cela coûtera cher. Pour cette raison, nous avons proposé une approche pour augmenter les données, sans perdre d'informations. D'autre part, le succès de l'architecture d'apprentissage en profondeur DL dans de nombreuses applications et l'extension du succès actuel des réseaux de neurones convolutifs (CNNs), dans le domaine de la classification et de l'analyse, a stimulé la proposition d'un nouveau système qui permet de classer les sons normaux à partir des sons

malades en convertissant le son en une image qui porte le graphique de la fréquence sonore, en fonction des paramètres MFCCs.

Notre projet se base sur la conversion du son en image passant vers les calculs de coefficients cepstraux de chaque son, et puis représenter ces matrices des coefficients dans des figures graphiques contenants des courbes illustratives. Le volume de la base de données est important pour réaliser ce type de projet et donc, au lieu d'une seule image pour chaque son, on va dissocier ces courbes dans plusieurs graphes. Pour l'exploitation des données disponibles concernant le but de l'étude, en gardant le même ordre de classification de ces courbes selon les fichiers du son d'origine. Pour cela nous avons discuté des concepts fondamentaux de l'apprentissage profond en général et des réseaux de neurones convolutifs en particulier en introduisant différents types de classes utilisées dans la classification ainsi que les méthodes d'organisation.

Le but de ce travail, est d'appliquer des techniques d'apprentissage en profondeur pour classer une image vocale pathologique, surtout pour identifier la maladie de Parkinson afin de fournir une aide au diagnostic des patients atteints de la maladie de Parkinson.

Ce mémoire se compose de quatre chapitres et s'organise comme suit:

- Dans le premier chapitre, nous présenterons l'appareil phonatoire et ses organes.
  - Puis les différents troubles de la voix décrire.
  - La maladie de Parkinson est ensuite discutée spécifiquement avec un bref historique, les causes, les symptômes ainsi que ses méthodes d'évaluation.
- Le deuxième chapitre s'intéresse à la présentation de l'apprentissage profond, dans lequel nous détaillons plus les réseaux de neurones convolutifs. Notre approche est présentée
- Dans le chapitre 3, et le dernier chapitre est consacré à la description des différents outils utilisés dans le développement de notre travail, ainsi que des différents résultats obtenus et leurs discussions.

Chapitre I

La Production De La Parole

Table des matières

1. Introduction 19
2. Anatomie de la parole 19
2.1 Voix humaine 19
2.2 Production de la voix 19
2.3 Appareil phonatoire humain 20
2.3.1 Appareil respiratoire 21
2.3.2 Larynx 22
2.3.2.1 Définition (généralité) 22
2.3.2.2 Structure des cordes vocales 23
2.4 Cavité résonante 24
2.4.1 Le pharynx 25
2.4.2 Cavité buccale 25
2.4.3 Cavité labiale 26
2.4.4 Fosse nasale (Cavité nasale) 26
2.4.5 Cavité orale 27
3.Pathologie de la parole : Dysarthrie Parkinsonienne 27
3.1 Introduction sur la maladie de Parkinson 28
3.2 Historique 30
3.3 Dysarthrie Parkinsonienne 31
3.3.1 Causes de la maladie Parkinson 31
3.3.2 Symptômes de parkinson 32
3.3.3 Diagnostic de Parkinson 33
4. Les méthodes d'évaluation d'une dysarthrie 33
4.1 Preceptive 33
4.2 Objective 34
5. Conclusion 35

Liste des Figures

Figure I.1 : Schéma Des organes principaux impliqués dans la production de la parole.....20
Figure I.2 : schéma de l'appareil phonatoire 21
Figure I.3 : Explication de l'appareil respiratoire humain.....21
Figure I.4 : Coupe sagittale de la bouche et du cou, comprenant le larynx et l'appareil supra-laryngé 22
Figure I.5 : schémas de coupe du larynx. 22
Figure .I.6: structure de larynx.....22
Figure I.7 : les positions des cordes vocales. 23
Figure I.8 : cavité buccale.....26
Figure I.9 : la cavité nasale 26
Figure I.10 : cavité oral.....27
Figure I.11 : Maladie de Parkinson 29
Figure I.12 : diminution de la libération de dopamine dans le cas MP 32

Liste des Acronymes

AO : Âges D'apparition

MP : Maladie de Parkinson

LDOPA : La lévodopa médicamenteuse antiparkinsonienne

1. Introduction

Ce chapitre résume l'anatomie de la voix humaine et le processus de la production de la parole. La première section de ce chapitre décrit le rôle de chaque composant de l'appareil phonatoire, puis, la pathologie de la parole concernant la dysarthrie Parkinsonienne est abordée. La dernière section de ce chapitre présente les méthodes d'évaluation parmi les plus utilisées, de la discrimination automatique d'une dysarthrie.

2. Anatomie de la parole

2.1 Voix humaine

La voix humaine est le son le plus important de notre environnement. La voix joue un rôle fondamental dans nos interactions sociales. Elle est tout d'abord le support du langage articulé dont nous devons comprendre le sens. Ainsi, il est aisé de comprendre que notre espèce a développé une expertise considérable dans le domaine de la perception de la voix, mais en plus de la perception du langage, nous sommes capables d'extraire de la voix de nombreuses informations socialement pertinentes.(TUBACH et al., 1990) À partir de la voix, nous sommes capables de déterminer approximativement le genre, l'âge, la taille, l'état affectif et surtout l'identité du locuteur.

2.2 Production de la voix

La production de la parole commence avec la formulation de la pensée à être communiquée. La personne qui parle, suite à des processus neurologiques et musculaires, produit les fluctuations de la pression de l'air qui constituent le signal vocal.(Combescure et al., 1995a)

Les caractéristiques acoustiques de la voix sont fondées sur ce qui détermine sa production, telle que la production de la voix. Cette dernière implique l'interaction d'une source et d'un filtre et il faut ajouter l'appareil respiratoire qui forme avec les poumons un réservoir d'air apportant l'énergie nécessaire à la production de sonore.

Selon la source et le filtre l'appareil vocal agissent et interagissent pour générer une grande variété de sons. (Zantis, 2002)

Figure I.1 présente les organes actifs dans la production de la parole, qui se trouvent dans la disposition anatomique du tractus vocal (de bas en haut): les muscles du système respiratoire, le larynx, le pharynx, le palais mou, la langue, la mâchoire inférieure et les lèvres. Telle que les sons de la parole proviennent de phénomènes aérodynamiques et acoustiques, de l'air comprimé dans les poumons qui sont modulés par des contractions et divers changements dans la longueur du tractus vocal; Constriction laryngée, source de son, crico-pharynx, constrictions palatines et orales caractérisant les parties phonémiques des langues, consonnes et voyelles.

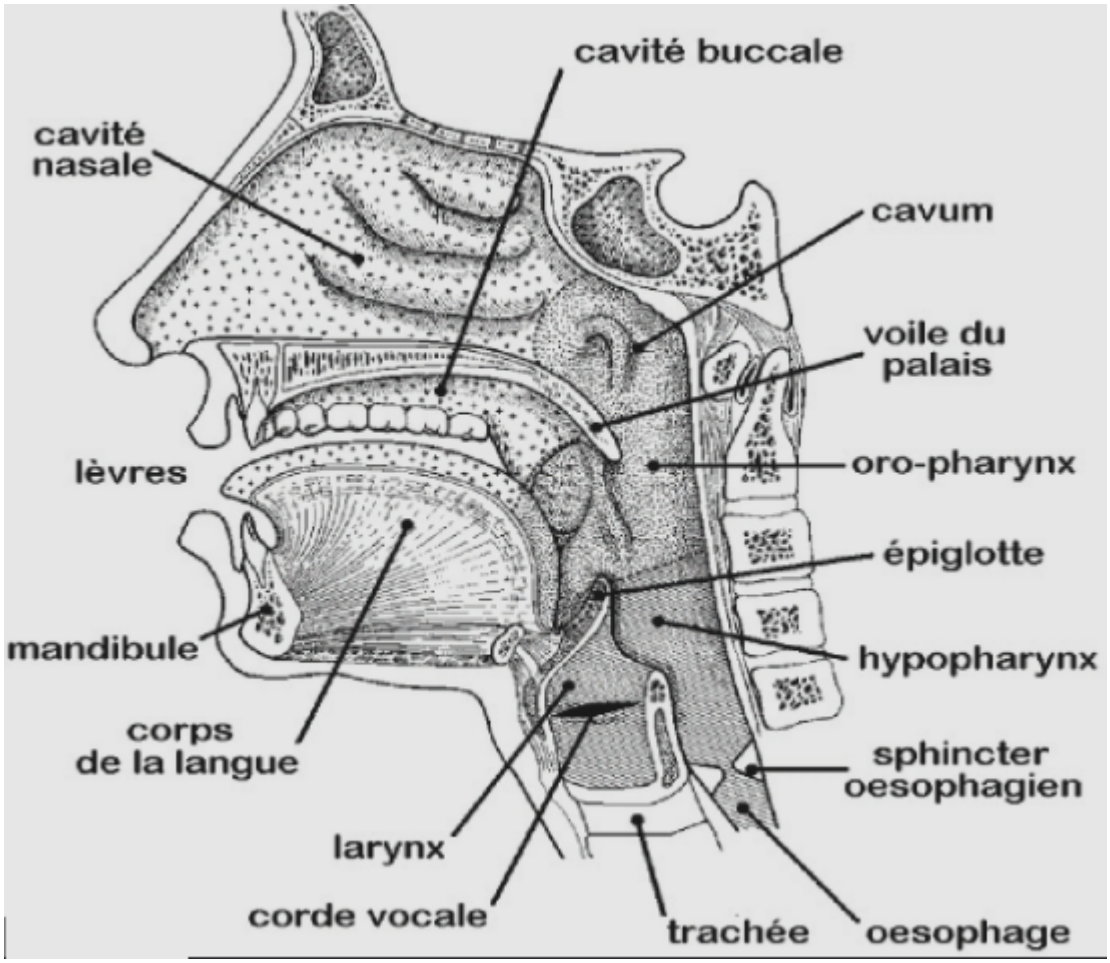


Figure I.1 : Schéma Des organes principaux impliqués dans la production de la parole

Le système de production de la voix humaine est divisé en trois parties principales: le système respiratoire, la source sonore et le filtre. Le système respiratoire se compose des deux poumons qui fournissent l'énergie nécessaire pour produire le son. La source du son est reliée à l'appareil du larynx, à travers les cordes vocales qui génèrent le son, et ce son est modifié par l'appareil au-dessus de la gorge, appelé filtre vocal. Le processus de respiration se déroule en deux étapes: inspirez et expirez. La contraction des muscles intercostaux et du diaphragme provoque l'expansion de la cavité thoracique et l'invocation de l'air. Lors d'une inhalation normale, les poumons reçoivent environ un demi-litre d'air oxygéné. Lorsque vous expirez, tous les muscles de la poitrine se détendent et le poids de la cage thoracique expulse passivement l'air. La production sonore apparaît principalement pendant la phase d'expiration. Habituellement, les phases d'inspiration et d'arrivée sont d'une durée relativement égale. Mais on peut observer que lors de la production de la parole, le cycle respiratoire change avec seulement 15% du cycle dédié à l'inspiration.(Massida et al., 2010)

2.3 Appareil phonatoire humain

La parole est une onde sonore issue du travail des structures anatomiques qui composent l'appareil phonatoire humain telle que ce processus coordonne par le système nerveux central à travers les étapes suivantes :

Les muscles abdominaux agissent sur le diaphragme et poussent l'air des poumons vers la trachée à son extrémité supérieure Le larynx régule le flux d'air sous forme d'impulsions périodiques appliquées au canal vocal. Il s'agit d'un ensemble de cavités : Cavité pharyngée suivie de la cavité buccale et contournant la cavité nasale. La luette qui prolonge le bord postérieur du palais mou aide à fermer la cavité nasale. Ainsi, lors de la production de la parole, la cavité nasale peut être entièrement ou partiellement conjuguée, voire séparée de la cavité buccale. D'autres dispositifs anatomiques importants qui sont impliqués dans la production de sons sont : la langue et les dents dans la cavité buccale et le nez dans la cavité nasale et les lèvres. (Rachedi et al., 2005)

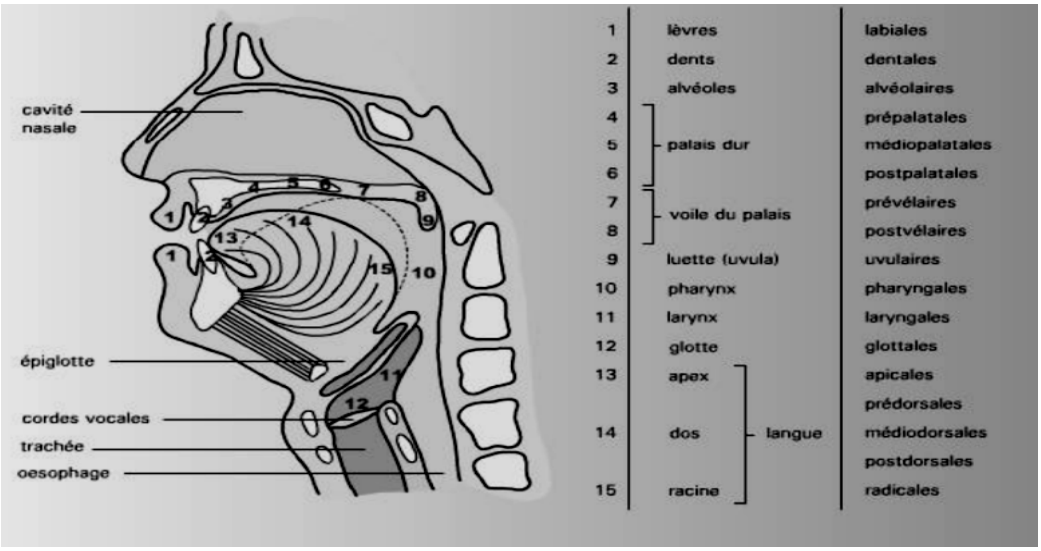


Figure I.2 : schéma de l'appareil phonatoire

2.3.1 Appareil respiratoire

La respiration humaine comporte deux étapes: l'inhalation et l'expiration. La contraction des muscles intercostaux et du diaphragme dilate la cage thoracique et appelle de l'air. Lors d'une inhalation normale, les poumons apportent environ un demi-litre d'air oxygéné. À l'expiration, tous les muscles de la poitrine se détendent et le poids de la cage thoracique expulse passivement l'air. La plupart de la production sonore se produit pendant la phase d'expiration.(Massida et al., 2010)

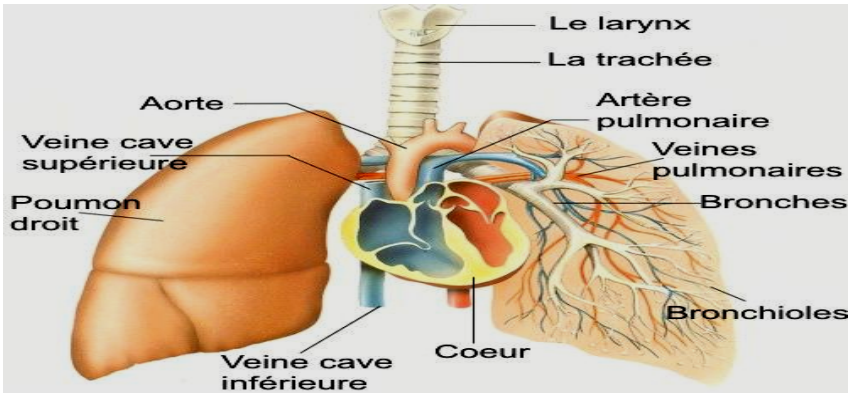


Figure I.3 : Explication de l'appareil respiratoire humain

2.3.2 Larynx

2.3.2.1 Définition (généralité)

Le larynx est un organe cartilagineux complexe du système respiratoire situé dans la gorge, devant le système respiratoire et la jonction digestive qui forme le pharynx. C'est le médiateur entre le pharynx et la trachée et comprend les cordes vocales. Le larynx a trois rôles principaux dans le corps: Premièrement, il contribue à la respiration en tant que partie intégrante du système respiratoire. Deuxièmement, il joue un rôle dans la déglutition car sa fermeture protège les voies respiratoires. Troisièmement, c'est un appareil pour produire des sons vocaux.(Drygajlo, 2014)

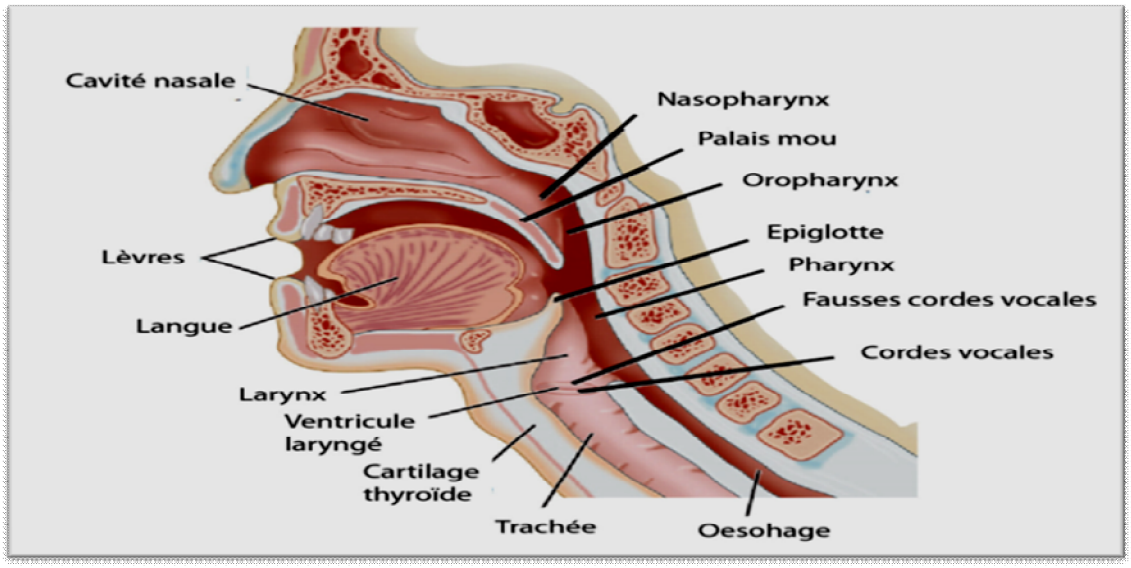


Figure I.4 : Coupe sagittale de la bouche et du cou, comprenant le larynx et l'appareil supra-laryngé

L'épiglotte est une structure cartilagineuse du larynx. Il a la forme d'une feuille dont la branche antérieure s'insère dans la fente du cartilage thyroïdien. L'extrémité postérieure est tournée vers le haut à l'état basal, permettant la respiration en permettant à l'air de circuler entre le pharynx et le larynx.(Rhonda et al., 2000)

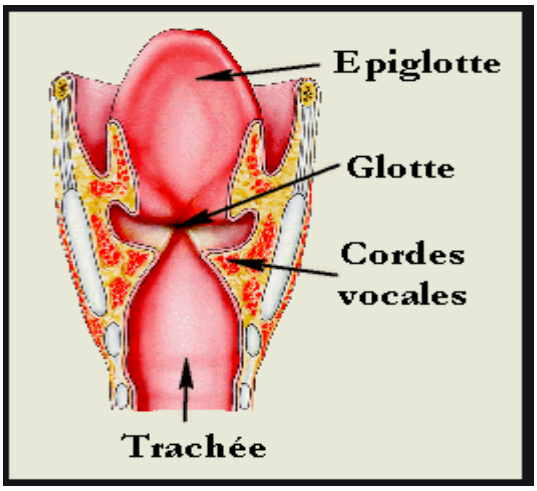


Figure I.5 : schéma de coupe du larynx.

Le larynx est formé d'un ensemble de muscles et cartilages mobiles entourant une cavité située à la partie supérieure de la trachée. Les cordes vocales, partie intégrale du larynx, peuvent le fermer ou peuvent former une ouverture variable appelée glotte. La fonction du larynx est de fournir une excitation périodique au reste du système sous la forme d'une suite d'impulsions périodiques de pression d'air pendant la phonation du son voisé. (Massida et al., 2010)

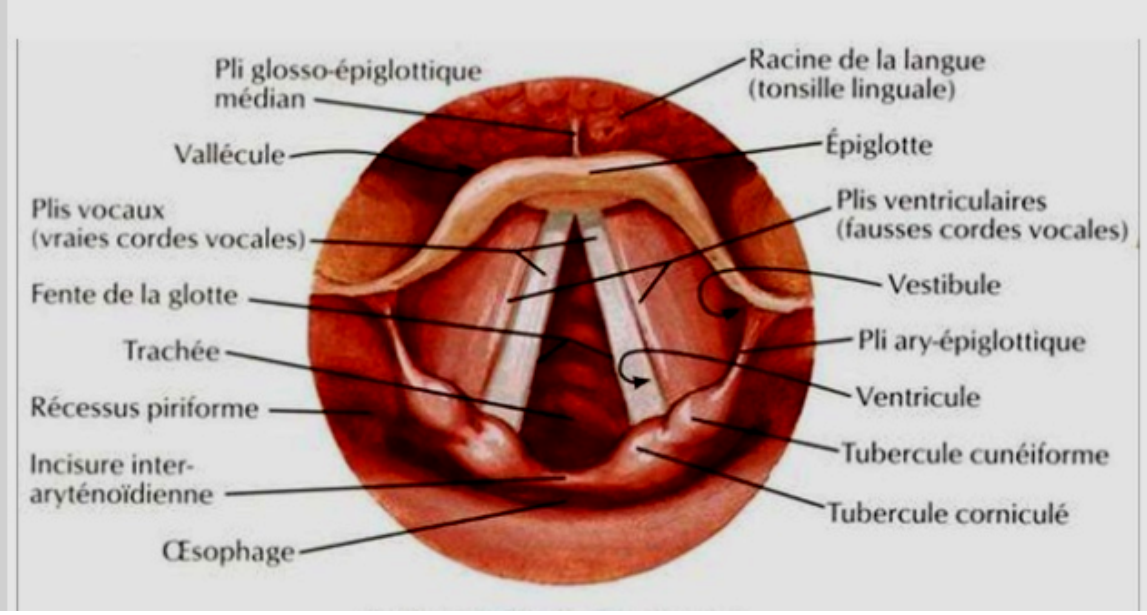


Figure .I.6 : structure de larynx

2.3.2.2 Structure des cordes vocales

Les cordes vocales sont les composants fonctionnels les plus importants de l'orgue vocal, car ils fonctionnent comme un générateur de sons vocaux. Ils sont recouverts de muqueuse et l'espace entre eux porte le nom de glotte, dont un point final se trouve sur le site de la pomme d'Adam. Des images de la glotte humaine sont présentées dans la figure suivante :

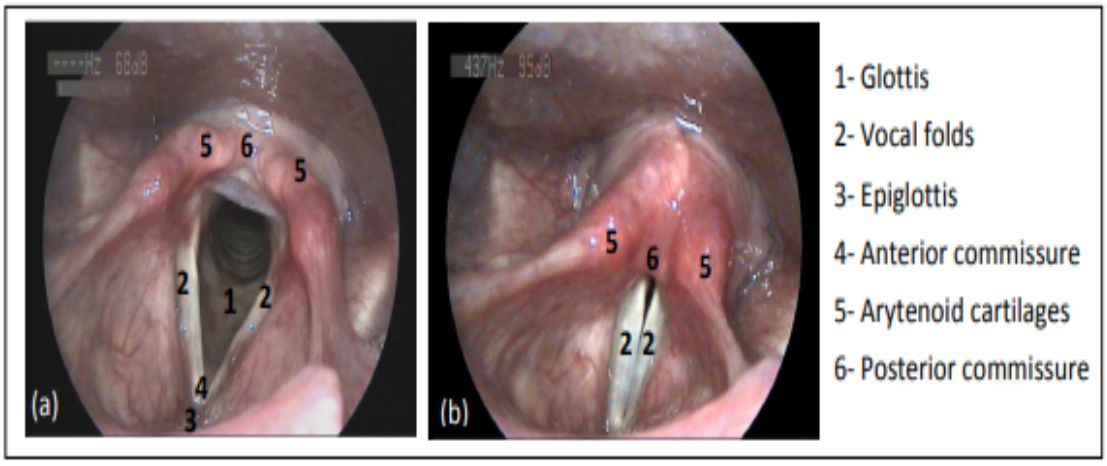


Figure I.7 : les positions des cordes vocales

La longueur des cordes vocales varie : chez un nouveau-né, elle est d'environ 3 mm et augmente à 9-13 mm - et 15-20 mm - respectivement chez la femelle adulte et le mâle.

(Sundberg, 1954). Cette longueur est ce qui définit la fréquence de vibration des cordes vocales, appelée fréquence fondamentale (ou hauteur). La fréquence étant inversement proportionnelle à la longueur, les valeurs de la fréquence fondamentale des voix féminines sont supérieures à celles des voix masculines. Pour les voix féminines, les valeurs de la fréquence fondamentale sont proches de 220 Hz tandis que pour les voix masculines sont d'environ 110 Hz. Ainsi, lorsqu'un ténor chante une note avec une fréquence fondamentale de 330 Hz, cela signifie que ses cordes vocales s'ouvrent et se ferment 330 fois par seconde. Les cartilages aryténoïdes contrôlent le mouvement des cordes vocales, les séparent, en cas de respiration, et les joignent et les resserrent pour produire une émission sonore voisée. L'action du cartilage joignant les cordes vocales est connue sous le nom d'adduction et l'action inverse est l'abduction. C'est la combinaison des actions d'adduction et d'abduction, lorsqu'elles sont effectuées à certaines fréquences, qui provoquent la production d'une onde sonore et qui se propage ensuite vers l'ouverture des lèvres. Les sons produits à la suite de la vibration des cordes vocales sont appelés sons vocaux, et les sons produits sans vibration de ces dernières (les cordes vocales restent ouvertes uniquement et l'excitation glottale est bruyante) sont appelés non vocaux. (Sundberg, 1954).

Le tube formé par le larynx, le pharynx et les cavités buccales et nasales, est appelé le tractus vocal, il peut donc être défini comme l'espace en aval de la glotte qui se termine par la cavité buccale ou les narines. La morphologie individuelle détermine sa longueur, mais chez les mâles adultes, la longueur du tractus vocal est d'environ 17 à 20 cm et 3 cm de diamètre. Les enfants et les femelles adultes ont des voies vocales plus courtes. On sait que plus le conduit vocal n'est long, plus les fréquences des formants sont basses.(Dias, 2012)

Lorsque les cordes vocales vibrent et forment des impulsions de pression près de la glotte (qui, à leur tour, se propagent vers les ouvertures vocales et nasales), l'énergie des fréquences de l'excitation est modifiée lorsque celles-ci se déplacent dans le tractus vocal. Par conséquent, le son produit est façonné par les cavités résonantes , au-dessus de la source glottale et, par conséquent, le tractus vocal est responsable du changement acoustique de la source vocale.(Zantis, 2002)

2.4 Cavité résonante

La cavité résonnante est un espace creux à l'intérieur d'un solide dans lequel une onde entre en résonance. Cette onde peut être sonore ou lumineuse. Pour l'être humain, chaque passage de l'air se fait à partir des résonateurs de l'appareil vocal. Ces résonateurs sont :

- Le pharynx
- La cavité buccale
- La cavité labiale
- La fosse nasale

Chaque cavité a une fréquence de résonance spécifique qui dépend de ses dimensions, du matériau qui la compose et du type d'onde. Telle que chaque partie du résonateur a une propre fréquence de résonance.(Combescure et al., 1995b) Les principes composants de tractus vocal est pharynx, la cavité nasale, la cavité buccale et la cavité labiale.

2.4.1 Le pharynx

Le pharynx est un conduit musculo-membraneux étendu de la base du crâne en haut à la sixième vertèbre cervicale en bas. Cet organe fondamental est unique, commun aux voies respiratoires supérieures et à la partie initiale de l'digestif. Le pharynx a la forme d'un entonnoir irrégulier dont la longueur et le diamètre peuvent varier en fonction de l'activité des muscles qui le constituent et il intervient dans la respiration, la déglutition, la phonation, la gustation et l'olfaction. (Sundberg, 1954) Il peut être divisé en 3 parties, le laryngopharynx, l'oropharynx et le naso-pharynx.

- **Le laryngopharynx :** est le respiratoire et digestif dans sa partie supérieure, et il est uniquement digestif dans sa partie inférieure, où il se prolonge avec l'œsophage. Il se termine en bas au niveau de la bouche de l'œsophage. Sa hauteur totale est de 14 à 15 cm au repos ; lors de la déglutition, son extrémité inférieure s'élève d'environ 3 cm, accompagnée du bloc laryngo-trachéal et du corps thyroïde. Il constitue la partie cervicale du pharynx. [1]
- **L'oropharynx :** c'est la partie orale, et également céphalique du pharynx. C'est une partie respiratoire et digestive, elle est assez étendue et béante. Elle environs de 5 cm de largeur et 5cm de profondeur.[2]
- **Le naso-pharynx :** cette partie du pharynx est respiratoire et phonatoire, c'est une partie toujours béante, où l'aire circule librement. Le naso-pharynx s'étend depuis la base du crâne en haut, jusqu'à la face supérieure du palet mou en bas.

2.4.2 La cavité buccale

C'est la partie initiale de l'appareil digestif, contient l'organe de la gustation. Destinée à la mastication et à l'insalivation des aliments et joue un rôle important dans la phonation, elle est subdivisée en deux parties par les arcades alvéolo-dentaire :

- La cavité périphérique répondant au vestibule de la bouche : c'est un espace en forme de fer à cheval compris entre les lèvres et les joues en avant, et les arcades alvéolo-dentaires en arrière. [3]
- La cavité buccale proprement dite : elle est limitée sur son pour tour par des arcades alvéolo-dentaire en avant et sur les côtes, en haut par la voûte palatine et le palais mou, en arrière par l'isthme du gosier. (Buccale, 2015)

[1] Article d'anatomie du pharynx vue de profil, Collège francais d'ORL DCEM (2009) pg 1

[2] Article de R.MARIANOWSKI ELSEVIER, Chronic Pharyngites Introduction Service d'otorhinolaryngologie, centre hospitalier Morvan, Brest France 2005

[3] Article de N. Perone et J-P Humair. Revue Médicale Suisse, Diagnostic et prise en charge de la pharyngite de l'adulte, 2009. Laboratoire d'anatomie de médecine Enseignements.pg 7

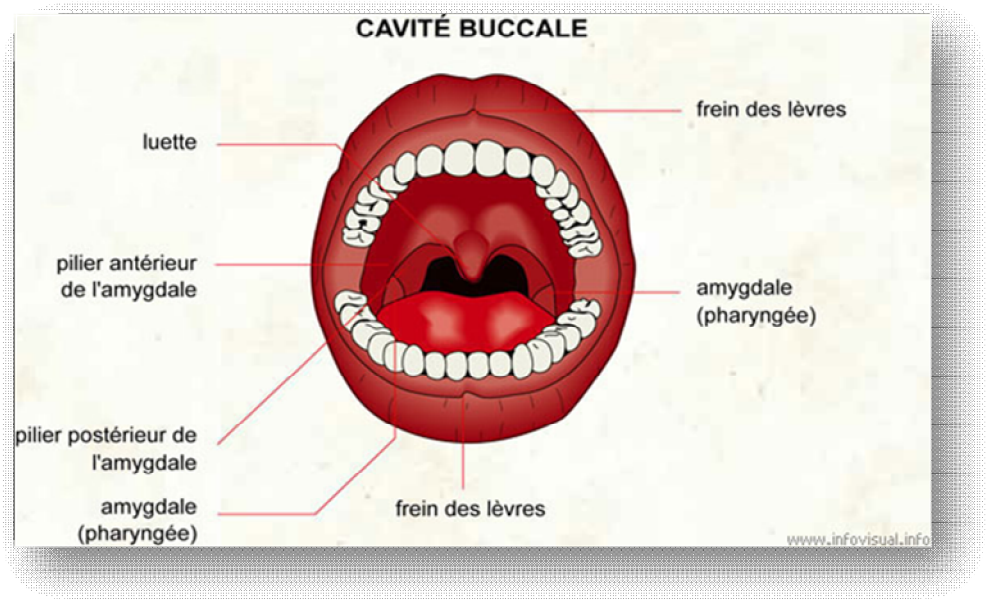


Figure I.7 : cavité buccale

2.4.3 Cavité labiale

C'est une cavité qui se forme lorsque les lèvres sont projetées vers l'avant, à travers des plis intermédiaires de la membrane muqueuse qui relient la surface interne des lèvres aux gencives. Et c'est le siège principal de l'expression et de la résonance. Le muscle primaire de la lèvre circulaire assure à son tour l'occlusion des lèvres. Et cela permet aux lèvres d'avancer pour faire avancer les lèvres et faire un blocage aux lèvres et les rapprocher. Les différents organes constitutifs jouent un rôle dans l'expression à la fois des voyelles et des consonnes: lèvres, dents, langue et mâchoire. (Massida et al., 2010)

2.4.4 Fosse nasale (La cavité nasale)

C'est la position du palais mou qui appartient à la cavité buccale et déterminer si l'air expiré durant la phonation s'échappe uniquement par la bouche (palais mou relevé), et le nez (palais mou partiellement abaissé) ou uniquement par le nez (palais mou totalement abaissé). La cavité nasale est donc peu mobile, mais le passage de l'air par cette cavité modifie considérablement la qualité du son du résonateur.(Teston, 2007)

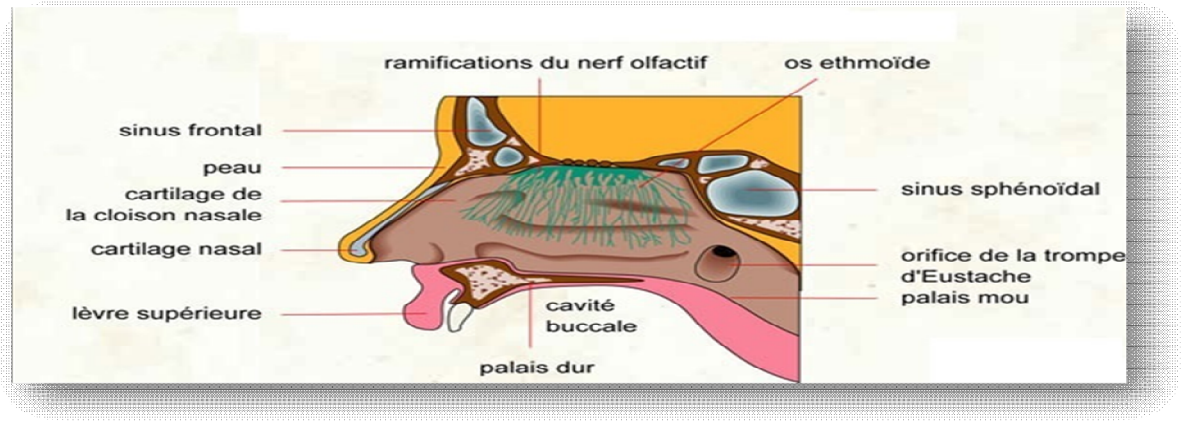


Figure I.8 : la cavité nasale

2.4.5 La cavité orale

Il est situé devant et à côté des dents et derrière l'isthme de la gorge. Il comprend les dents, les joues, la langue, la luette, les amygdales et le palais mou. composer de 17 muscles. Son rôle principal est de faire entendre des voix. Au repos, les lèvres sont en contact les unes avec les autres. Dans la parole, des mouvements complexes sont produits qui modifient la longueur du résonateur buccal et le degré d'ouverture.(Laprie, 2000)

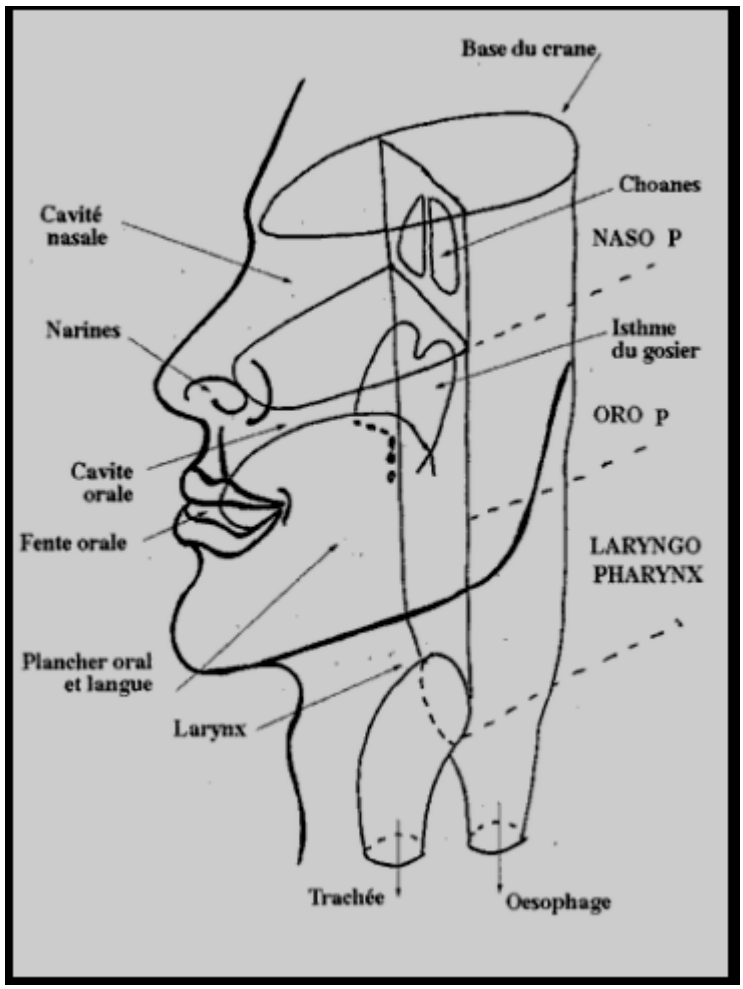


Figure I.9 : cavité oral. [4]

3. Pathologie de la parole : Dysarthrie Parkinsonienne

Les pathologies de l'appareil vocal sont multiples et sont à l'origine d'importants handicaps, sur le plan physique, comme sur le plan social. Parmi les maladies les plus répandues, sont celles relatives à l'appareil vibrateur, constitué principalement du larynx et des cordes vocales.(TUBACH et al., 1990)

Et convoquer un problème de communication lié à la parole. Il s'agit d'un trouble caractérisé par l'impossibilité d'émettre un son articulé et modulé dont l'enchaînement des syllabes constitue des mots compréhensibles. Il y a deux types de pathologie organique et fonctionnaire.

[4] http://d.lefebvre.pagesperso-orange.fr/cavite_orale.htm

La pathologie organique c'est une dysphonies d'origine organique morphologique regroupent les altérations vocales liées à une modification anatomique des cordes vocales ou d'une partie du larynx telle que nodules, polypes, œdèmes de Reinke, kystes, encoches, sulcus, tumeurs...etc. (Teston, 2007). Les facteurs organiques qui affectent la voix et l'articulation (audition, expression vocale, articulation ...) et les perturbations de chacun de ces facteurs peuvent être à l'origine de troubles de la parole. [5] La dysphonie est intéressante dans la mesure où les dysphonies d'origine organique périphérique correspondent à une atteinte biomécanique localisée en fin de parcours moteur et à l'origine de la source vocale. (Ghio, 2007)

Par exemple des organes du système nerveux, tels que:

- Fente palatine: Le palais est divisé en deux types: le palais dur et le palais mou, de sorte que tout défaut de celui-ci entraîne des troubles de la parole.

- Fente labiale: Les lèvres sont deux éléments importants dans le processus d'influence sur la qualité et le type de voix.

- Problèmes de la langue: c'est l'organe le plus important dans la production de la parole, et parmi les problèmes qui lui surviennent et qui affectent la parole sont les suivants: les nœuds de langue, Différence de taille de langue, Tumeurs de la langue...

La pathologie fonctionnelle : Le concept de dysphonie fonctionnelle ou dysfonctionnelle est équivalent. L'adjectif « dysfonctionnel » nous paraît préférable dans la mesure où il traduit plus précisément le dysfonctionnement et l'utilisation défaillante de l'appareil phonatoire. (Moerman et al., 2006) définit les dysphonies dysfonctionnelles comme « le résultat d'une inadéquation du comportement vocal responsable d'une diminution de l'efficacité du geste de la parole ». Pour Remacle (1999), il s'agit d'une « altération du geste vocal en absence de perturbation organique permanente à l'origine de cette dysphonie ». Cela se traduit essentiellement par un usage inapproprié des tensions musculaires, soit en hypertonie, soit en hypotonie. L'hypertonie est problématique dans la mesure où elle est génératrice de malmenage vocal. Ce forçage peut induire des modifications physiologiques qui entraînent, généralement, une baisse de rendement laryngé. Cette diminution d'efficacité favorise alors une compensation sous forme d'accroissement des tensions et le locuteur entre ainsi dans un mouvement pouvant aboutir finalement à une lésion organique. On assiste alors à une forme complexe de dysphonie.[6]

3.1 Introduction sur la maladie de Parkinson

La maladie de Parkinson a été décrite pour la première fois en 1817 par Sir James Parkinson. Elle est, après la maladie d'Alzheimer, la deuxième maladie neurodégénérative la plus répandue dans le monde et, après les accidents vasculaires cérébraux, la deuxième cause

[5] Article L'intrication entre symptômes organiques et symptômes fonctionnels : une source d'affolement Khadija Lahlou-La forêt Dans Champ psychosomatique publier le janvier 2009.
[6] Article Troubles du contrôle moteur de la parole : contribution de l'étude des dysarthries et dysphonies à la compréhension de la parole normale. Des auteurs Serge Pinto, Alain Ghio Dans Revue française de linguistique appliquée publié le février 2008.

de handicap moteur chez les sujets âgés. En France, elle touche plus de 100 000 personnes avec une incidence d'environ 8000 nouveaux cas par an.(Pinto et al., 2010)

Fin 2015, 160 000 patients parkinsoniens traités, avec environ 25 000 nouveaux cas par an 17% des nouveaux cas étaient âgés de moins de 65 ans. Les hommes sont atteints 1,5 fois plus que les femmes. Augmentation importante du nombre de cas en raison du vieillissement de la population et de l'amélioration de l'espérance de vie. En 2030, le nombre de patients parkinsoniens pourrait augmenter de 56% avec 1 personne sur 120 de plus de 45 ans atteinte de la maladie. Parmi les 527 423 décès survenus chez les sujets âgés de 50 ans et plus en 2014, 1,8% mentionnaient la maladie de Parkinson (soit 9 638 décès).[7]

Les personnes décédées de maladie de Parkinson sont en moyenne plus âgées que les personnes qui ne décèdent pas de cette maladie (84 ans vs 80 ans), moins souvent des femmes, plus souvent mariées et deux fois plus à risque de décéder en maison de retraite.L'anxiété et la dépression deviennent plus courantes chez les personnes ayant subi la maladie. En termes de facteurs liés au sexe, le mâle a la possibilité de contracter la MP par rapport à la femelle dans un rapport d'environ 3: 2. Même s'il n'y a pas de remède pour cette maladie, certains traitements promettent une amélioration de la condition. La lévodopa médicamenteuse antiparkinsonienne (LDOPA) associée à des agonistes dopaminergiques montrent manifestement une réponse positive de la part des patients atteints de la MP. En cas de résistance aux médicaments dans les cas graves, une méthode chirurgicale efficace connue sous le nom de stimulation cérébrale profonde est effectuée pour réduire les symptômes moteurs. La MP peut également entraîner des troubles neuropsychiatriques allant de légers à graves. Ces types sont classés en fonction de la cognition, du comportement, de la pensée et de l'humeur. (Blond, 2010)

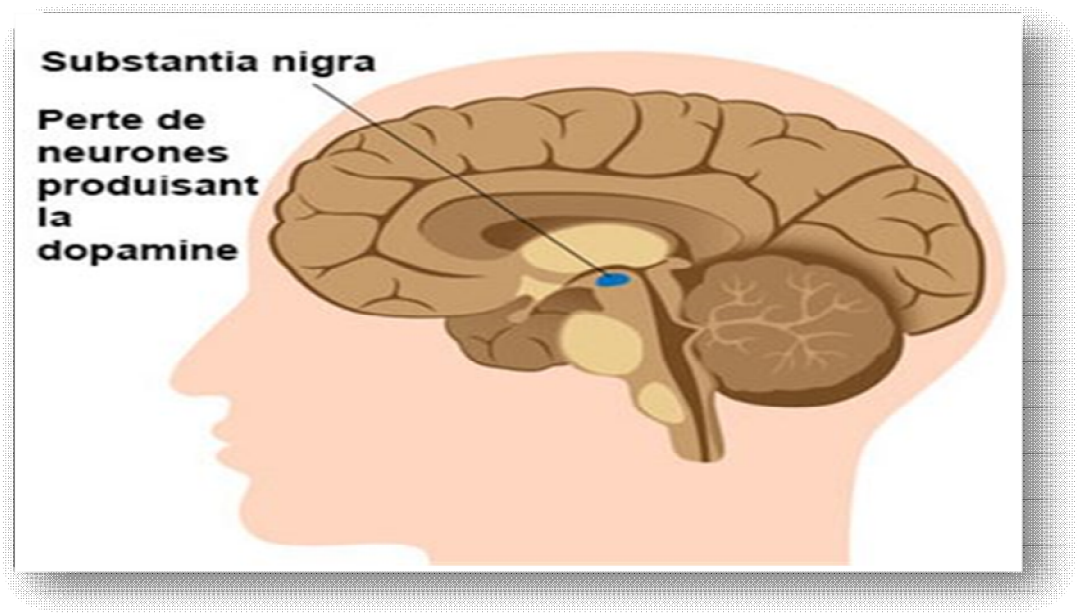


Figure I.10 : Maladie de Parkinson

[7]https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-liees-au-travail/maladie-de parkinson

3.2 Historique

Cette maladie a été décrite pour la première fois en 1817, par un médecin anglais dont elle a gardé le nom. Avec la publication d'un livre de 66 pages intitulé "An Essay on the Shaking Palsy". Sous la signature de James Parkinson lui-même. (Goetz, 2011). Il a défini la paralysie agitée : "mouvement de tremblement involontaire, associé à une diminution de la force musculaire, se produisant dans les parties du corps au repos (non en action) et même soutenu, avec une propension à plier le tronc vers l'avant et à passer de marcher au rythme de la course, les sens et l'intellect restant intacts "(Khalil, 1996).

La monographie de James Parkinson a ainsi ouvert la première étape de l'histoire de la maladie, la phase clinique.(Goetz, 2011).

On résume brièvement les dates importantes associées à cette maladie comme suit :

- En 1872, Jean-Martin Charcot, médecin et neurologue français .Lui et ses étudiants ont décrit avec diligence le spectre clinique de la maladie de Parkinson, citant deux modèles, le tremblement et la forme rigide / cognitive. Ils ont décrit en détail tous les changements dans l'arthrite, la dyskinésie et la douleur qui peuvent accompagner la maladie de Parkinson. Charcot a été l'un des premiers à approuver la suggestion d'utiliser le terme «maladie de Parkinson».

- En 1895, Richter et Maige ont fourni des détails cliniques et morphologiques sur les stades progressifs de la maladie de Parkinson, et ont été parmi les premiers à partager des dessins et des statues qui restent parmi les principaux documents picturaux de la maladie de Parkinson.

- En 1920, Brissot est le premier qui proposer sa suggestion de lésions de la matière noire comme siège anatomique de la maladie de Parkinson, et Treti'akoff et Foix et Nicolesco poursuivirent tout les études pathologiques du mésencéphale concernant la maladie.

- En 1950, deux découvertes séminales se produisent: localisation de la dopamine dans le cerveau, plus précisément dans le striatum; et le développement du modèle de la réserpine, qui sera plus tard utilisé comme premier modèle de parkinsonisme qui a été inversé par le traitement à la lévodopa.

- En 1950, deux découvertes fondamentales ont eu lieu: la localisation de la dopamine dans le cerveau, et plus précisément dans le striatum. Le développement du modèle de réserpine, qui sera par la suite utilisé comme premier modèle de la maladie de Parkinson réversible par la lévodopa.

- En 1953, l'analyse pathologique la plus complète de la maladie de Parkinson et l'identification claire des lésions du tronc cérébral ont été effectuées par Greenfield et Bosanquet.

- En 1960, Ehringer et Hornykiewicz se sont tournés vers le cerveau humain et, après avoir examiné une série d'échantillons de contrôle, ils ont découvert une déplétion fatale de dopamine dans la maladie de Parkinson et le cerveau post parképhalitique parkinsonisme.

- En 2008, les essais contrôlés par placebo sont devenus la norme dans la maladie de Parkinson, même dans le domaine de la chirurgie, principalement parce qu'une grande proportion de patients traités par placebo ont vu une amélioration objective de la maladie de Parkinson.

De nombreuses personnes célèbres sont porteuses de la maladie de Parkinson, comme le célèbre Muhammad Ali Cassius Clay, Harry Truman, Adolf Hitler, Joseph Fawzy, Robin Williams, Salvador Dali, Ronald Reagan, Michael J.Fox, Yasser Arafat ... etc. Et cela a été appelé maladie des célébrités.

3.3 Dysarthrie Parkinsonienne

3.3.1 Causes de la maladie Parkinson

La maladie de Parkinson est une maladie dégénérative caractérisée par une dégénérescence progressive, accélérée et prématurée des neurones dopaminergiques d'une région du tronc cérébral, la substance noire (pars compacta du locus niger). La dopamine est un neurotransmetteur qui intervient dans l'initiation, la commande, la prévision et l'exécution des gestes. Elle régule l'activité du locus niger qui traite les informations sensorielles, comportementales et émotionnelles. Les lésions de la pars compacta (produisant donc une dépopulation en neurones dopaminergiques) sont responsables d'une augmentation de l'activité des neurones glutaminergiques du noyau su thalamique, sur lesquels la dopamine a habituellement un rôle inhibiteur. Cela crée alors une hyperactivité du pallidum interne, qui entraîne à son tour une inhibition du thalamus moteur. Ce dernier régit l'activation des aires corticales motrices, mais la perturbation d'origine dopaminergique ne lui permet plus un contrôle normal, ce qui donne lieu à la symptomatologie parkinsonienne. La raréfaction neuronale atteint aussi d'autres noyaux du tronc cérébral non dopaminergiques, comme le locus coeruleus, le noyau dorsal du vague, la formation réticulée mésencéphalique, le noyau basal de Meynert (Petit & al. 1994), Rendant compte des signes de la maladie résistant au traitement dopaminergique (troubles axiaux, troubles cognitifs…).

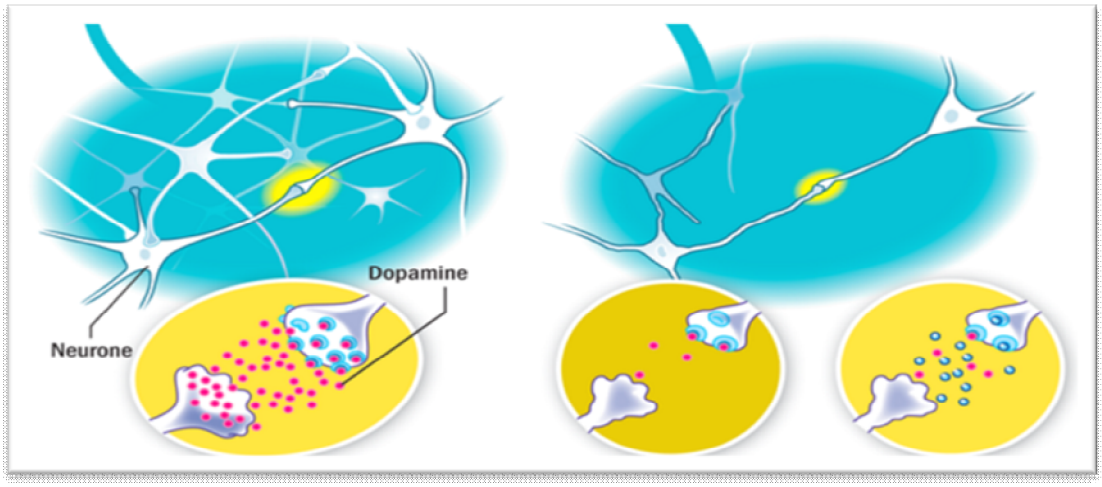


Figure I.11 : diminution de la libération de dopamine dans le cas MP

3.3.2 Symptômes de parkinson

Les études fait sur la maladie parkinson dit que les symptômes de PD commencent généralement progressivement et s'aggravent avec le temps. Au fur et à mesure que la maladie progresse, les personnes peuvent avoir des difficultés à marcher et à parler, des tremblements dans les mains, les bras, les jambes, la mâchoire ou la tête, la raideur des membres et du tronc, la lenteur des mouvements, un équilibre et une coordination altérés, entraînant parfois des chutes de difficulté à avaler. , mâcher et parler (troubles de la parole). Ils peuvent également avoir des changements mentaux et comportementaux, des problèmes de mémoire, des problèmes de sommeil, une dépression. (Blond, 2010)

Il existe de nombreuses études et recherches dans le domaine de l'étude de la maladie de Parkinson, et de nombreux symptômes ont été découverts pour confirmer la maladie. Nous résumons les symptômes les plus importants dans les points suivants:

- Au niveau central du cerveau, divers symptômes apparaissent, tels que dépression, anxiété, diminution des performances intellectuelles et le patient devient étourdi (ou s'évanouit dans certains cas).

- La personne souffre également de fermeture forcée des paupières au niveau des yeux, de difficultés à parler par excrétion des lettres avec salivation excessive. Difficulté à avaler et la voix du patient passe de douce à basse.

- Il existe également d'autres symptômes au niveau des poumons: obstruction des voies respiratoires supérieures, anomalies du contrôle de la vessie.

- Le patient souffre également d'une transpiration accrue, d'une faiblesse générale du corps avec une sensation de douleur, de progressivement les tremblements deviennent visibles à l'examen clinique au repos des membres. Aussi, de léthargie, d'un équilibre variable, d'une petite écriture, de somnolence, une diminution du ballant des bras lors de la marche, et de difficultés à bouger ou à marcher, en particulier aux stades avancés de la maladie.

Cette maladie avance cachée c'est pour cela qu'elle est difficile à étudie. La recherche mondiale et d'autres thérapies sont en cours de développement tels que les thérapies cellulaires, géniques, et immunologique.

3.3.3 Diagnostic de Parkinson

De nombreuses personnes ignorent les premiers symptômes de la maladie de Parkinson ont des effets similaires au vieillissement normal à ses débuts. Il n'y a pas des tests médicaux pour la détection finale de cette maladie, il peut donc être difficile de diagnostiquer avec précision la maladie de Parkinson. De plus, de nombreux patients souffrent d'une détérioration de leur situation financière, en raison des frais médicaux qui y sont associés et de la difficulté à se déplacer et à parcourir de longues distances pour diagnostiquer la maladie. Heureusement, depuis la découverte du traitement du signal et l'étude des images analytiques, des experts avancés ont utilisé des mesures discriminatoires de la voix des personnes atteintes de maladie ou la détection de la MP. En raison de l'apparition précoce de troubles des cordes vocales dans la maladie de Parkinson, il peut être détecté à un stade précoce de la maladie, ce qui signifie qu'il peut être identifié en analysant les données vocales du patient. Pour détecter cette maladie, il est plus facile, meilleur et plus économique d'utiliser la détection des maladies avec apprentissage en profondeur en raison de son développement et de son épanouissement. (Camps et al., 2018)

4. Les méthodes d'évaluation d'une dysarthrie

4.1 Perceptive

Depuis la première grille clinique perceptive de (Darley et al. 1969). L'évaluation de la dysarthrie s'articule autour de différentes approches complémentaires. La plus immédiate en pratique clinique est l'analyse perceptive. Cette dernière consiste à relever les anomalies rencontrées lors de l'écoute de la parole du patient. En 2001, Parais et Auzou élaborent une grille française d'analyse perceptive sur 32 critères répartis en 8 catégories. Cette étape de l'évaluation est indispensable à la pose du diagnostic, « un patient est dysarthrique parce qu'il est perceptivement reconnu comme tel. » (Duffy (2005), cité par Auzou, 2007).

La réduction de l'intelligibilité est un des critères principaux de l'évaluation clinique de la dysarthrie (Amosse et al., 2004). Ce phénomène complexe qu'est l'intelligibilité est défini par Yorkston & Beukelman (1980, cité par Auzou et al., 2006) comme le degré de précision avec lequel le message est compris par l'auditeur. Apprécier l'intelligibilité d'une personne dysarthrique implique de déterminer à la fois la réalisation acoustique perturbée dans sa parole, mais également les stratégies mises en place par l'auditeur pour le comprendre. Toutefois, l'intelligibilité est à distinguer d'abord de la compréhensibilité, plutôt présente en contexte de communication, à savoir « le degré avec lequel un auditeur comprend la parole à partir du signal acoustique (intelligibilité) et des autres informations qui contribuent à la compréhension de l'énoncé » (Auzou, 2007) ; et ensuite de l'efficacité, la quantité de message comprise par unité de temps. La perte d'intelligibilité, plainte principale des patients et de leur entourage, constitue également un indice de sévérité de la dysarthrie, et reflète donc le handicap communicationnel du patient. L'évaluation de l'intelligibilité est donc nécessaire.

Elle peut s'effectuer sur diverses productions verbales : phonèmes, mots, phrases, lecture à voix haute de texte, conversation libre. Dans l'évaluation Clinique de la Dysarthrie (Auzou et al. 1998) devenue Batterie d'évaluation Clinique de la Dysarthrie (BECD) (Auzou et al. 2006), adaptations du Frenchay Dysarthria Assesment d'Enderby (1983), le Score d'Intelligibilité (SI), obtenu grâce à une analyse perceptivo-phonétique de mots, de phrases et en conversation, permet d'attester la présence d'une dysarthrie (avec un score ≤ 23) et d'en définir la sévérité. Un SI compris entre 18 et 23 indique une dysarthrie légère ; si le SI se situe entre 12 et 17, il s'agit d'une dysarthrie modérée ; une dysarthrie sévère est signalée par un SI compris entre 7 et 11. Un SI équivalent à 6 et moins rend compte d'une dysarthrie très sévère (ou massive) avec une parole inintelligible.

4.2 Objective

Le son de parole ayant pour origine des processus aérodynamiques il en a été de même pour des évaluations articulatoires à travers l'étude de ces phénomènes. Mais ces techniques analytiques sont mal adaptées à une utilisation clinique de routine Depuis peu de temps on voit apparaître des applications spécifiques à l'évaluation clinique, pour la prosodie (c'est l'étude des traits phoniques, c'est-à-dire l'étude du rythme (vitesse d'élocution), de l'accent et de l'intonation.) et les fuites nasales en particulier. Ces développements permettent d'envisager à l'avenir une diffusion plus générale des évaluations objectives des dysarthries.

- **Les méthodes aérodynamiques**

Le domaine de l'aérodynamique dans la production de la parole est fondamental par le fait qu'il est à l'origine de toutes ses manifestations sonores. Ces méthodes instrumentales d'investigation de la parole sont les plus anciennes, Les paramètres aérodynamiques sont constitués par les débits à la bouche (oral) et aux narines (nasal) et les pressions intra-orale et sous glottique. L'étude des paramètres aérodynamiques donne des indications très riches sur la réalisation des consonnes, tant sur leur dynamique que sur leur synchronie par l'étude du débit d'air buccal pour les occlusives et constrictives et du débit d'air nasal pour l'évaluation du déficit vélaire. Ces méthodes, dont la pratique n'est pas très compliquée, nécessitent une bonne connaissance des mécanismes articulatoires et se prêtent mal à une investigation automatique, si ce n'est pour des mesures globales de fuites vélaires par exemple. Les systèmes de mesures aérodynamiques ne sont pas très nombreux. On peut citer le masque de Rothemberg très diffusé aux USA et l'Aérophone de Frokjer Jensen commercialisé par Kay Elemetric. Ces systèmes sont entachés de plusieurs défauts. Seul le système EVA proposé par la société SQLab permet des mesures nasales et une faible gêne articulatoire.(Pinto et al., 2010)

- **Les méthodes acoustiques**

Les méthodes d'investigation acoustique sont basées sur l'étude du signal de parole selon les deux aspects de l'information linguistique, segmentale ou suprasegmentale. Dans le premier cas, on étudie particulièrement la durée des segments phonétiques, leur timbre (Formants) et leur qualité vocale. Dans le second cas on étudie le rythme et le débit de la parole, son intensité et sa mélodie.(Teston, 2007)

L'étude du timbre est réalisée au moyen de l'analyse acoustique. Il s'agit de la décomposition d'un signal de parole en ses différentes composantes fréquentielles sous la forme d'un spectre. On distingue l'analyse en bande large, pour l'étude des formants de l'analyse en bande étroite pour l'étude plus fine des harmoniques. La durée du signal analysé sur un spectre étant inversement proportionnelle à la largeur de bande d'analyse, l'analyse en large bande est utilisée pour les phénomènes articulatoires rapides (bruits d'explosion de consonnes etc..). La représentation de l'évolution des spectres en fonction du temps s'appelle un sonagramme. Pendant longtemps le SONAGRAPHE a été le seul analyseur qui permettait des études spectrales et des mesures de durée sur le signal de parole On doit à Ilse Lehiste la première description de formes spectrale aberrantes de segments phonétiques produits par des patients neurologiques.(Teston, 2007).

La méthode MFCC (Mel-Frequency Cepstral Coefficients) est basée sur l'extraction des caractéristiques du signal vocal de son cepstre, modifiée selon l'échelle de Mel, une échelle qui cherche à reproduire la perception auditive humaine. (Works & Recognition, 1959)

Malgré leurs imperfections, personne ne peut contester les progrès dont ont bénéficiés les évaluations objectives des dysarthries ces dix dernières années. Les méthodes acoustiques ont gagné en efficacité et en simplicité d'interprétation. Les méthodes aérodynamiques ont également fait de nombreux progrès, en particulier au plan de la mise en œuvre.

5. Conclusion

Ce chapitre a permis dans un premier temps de présenter la complexité du processus de production de la parole nécessitant différents mécanismes neurologiques, phonatoires et articulatoires et l'intervention et la coordination de plusieurs organes de l'appareil phonatoire. Nous avons ensuite défini la Pathologie de la parole concernons la Dysarthrie Parkinsonienne et Les méthodes d'évaluation d'une dysarthrie Perceptive et Objective.

Chapitre 02

Analyse Et Discrimination Automatique

Table des matières

1.Introduction 41
2.Analyse acoustique et discrimination automatique 41
2.1 Introduction (généralité) 41
2.2 Paramètres MFCCs 42
2.2.1 Définition 42
2.2.2 Étapes de l'extraction des paramètres MFCCs 42
2.3 Discrimination automatique 46
2.3.1 Deep Learning 46
2.3.1.1 Définition(généralité) 46
2.3.1.2 Avènement du Deep Learning 46
2.3.1.3 Origine de Deep Learning 46
2.3.2 De Machine Learning au Deep Learning 46
2.3.3 Fonctionnement de Deep Learning 47
2.3.4 Exemple de Deep Learning 48
2.3.4.1 Conduite automatisé 48
2.3.4.2 Défense et aérospatial Deep Learning 48
2.3.4.3 Recherche médicale 48
2.3.4.4 Automatisation industrielle 49
2.3.4.5 Traduction 49
2.3.4.6 Reconnaissance vocale au niveau quasi humain 49
2.3.5 Les Algorithmes de Deep Learning 49
2.3.5.1 Réseau de neurones 50
2.3.5.2 Réseau Neuronal Convolutif (CNN) 50
2.3.6 Réseaux de neurones convolutifs pour la classification d'images 54
2.3.7 Apprentissage Par Transfert 55
2.3.7.1 Modèle VGG16 56
2.3.7.2 Modèle Xception 56
2.3.7.3 Modèle Inception V3 57
2.3.7.4 Modèle ResNet 57
3. Mesures d'évaluation des modèles 58

4. Classification Des Images ..... **Erreur ! Signet non défini.**
5. Travaux récents utilisant Deep Learning pour la parole 61
6.Conclusion 62

Liste des Figures

Figure II.1 : Étapes d'extraction des coefficients MFCC 42
Figure II.2 : Hamming Window 44
Figure II.3 : Les filtres triangulaires passe-bande en Mel-Fréquence 44
Figure II.4 : Réseau de neurone ET Réseau Deep Learning 47
Figure II.5 : Classification des méthodes de L'apprentissage profond 50
Figure II.6 : Schéma d'un réseau de neurones convolutionnel 51
Figure II.7 : Matrices de l'image input de pixels et filtre 52
Figure II.8 : la fonction convolutée. 52
Figure II.9 : la fonction ReLU 53
Figure II.10 : Exemplaire de Max Pooling 53
Figure II.11 : L'architecture de CNN 54
Figure II.12 : L'architecteur de modèle VGG16 56
Figure II.13 : L'architecteur de modèle Xception 56
Figure II.14 : L'architecteur de modèle Inception V3 57
Figure II.15 : L'architecteur de modèle ResNet 58
Figure II.16 : matrice de confusion 59
Figure II.17 : Interprétation graphique de la mesure ROC-AUC. 60

Liste des Acronymes

ANN: Artificial Neural Network

CNN: Convolutional Neural Network

DBM: Deep Boltzmann Machine

DBN: Deep Belief Network

DL: Deep Learning

DNN: Deep Neural Network

FC: Fully Connected

FCN: Fully Convolutional Network

FN : False Négatif

FP : False Positive

FTT: Fast Fourier Transform

GAN: Le Generative Adversarial Network

GPU: Graphics Processing Unit

LSTM: Long Short-Term Memory

MFCC: Mel-Frequency Cepstral Coefficients

ML: Machine Learning

MP: Maladie de Parkinson

NLP: Natural Language Processing

ReLU: Rectified Linear Unit

ROC: Receiver Operating Characteristic

RNN: Recurrent Neural Network

TN: True Negative

TP: True Positive

VAD: Voice Activity Detection

TNR : Taux Vrai Négatif

1. Introduction

Ce chapitre résume l'analyse acoustique basée sur les paramètres Les coefficients cepstraux à fréquence Mel (MFCCs), l'une des techniques d'extraction des fonctionnalités les plus populaires afin d'analyser notre voix. Nous expliquons ensuite les concepts fondamentaux des réseaux de neurones à action directe et les modèles profonds de base dans la littérature. Nous discutons sur les différents modèles de l'apprentissage profonds et les résultats obtenus sur ce domaine de la maladie parkinson.

2. Analyse acoustique et discrimination automatique

2.1 Introduction (généralité)

Dans le processus de communication parlée, le canal acoustique est le support essentiel porteur d'information. Mais il ne faut pas oublier qu'il ne s'agit que d'un support, que d'une projection résultant des processus cognitifs et des mouvements articulatoires du locuteur. Si l'objectif de l'analyse acoustique est de quantifier la «qualité» du résultat sonore pris comme élément principal de la communication parlée, notamment en testant, par exemple, l'intelligibilité ou plus largement le processus de transmission d'information, la méthode est directe. Si, par contre, par l'analyse acoustique, l'observateur cherche un marqueur sensible des mécanismes biomécaniques, articulatoires ou cognitifs, il faut rester prudent car le lien articulation/acoustique n'est pas une simple relation bijective.(Ghio, 2007)

L'analyse acoustique permet de faire des mesures, obtenir des indices, dégagés des tendances, d'une façon quantitative et objective. On l'oppose souvent au jugement perceptif, éminemment subjectif, variable d'un auditeur à l'autre ou changeant selon le contexte pour un même auditeur. Aussi, l'analyse acoustique s'avère finalement comme un moyen plus crédible d'évaluer et mesurer les dysfonctionnements de la voix et de la parole. Cependant, ces techniques ne sont pas non plus infaillibles car, comme tout procédé de mesure, elles sont sujettes à des erreurs, à des problèmes de non reproductibilité, à de la matière première inadaptée. Pour le moment, la plupart des études acoustiques sur les dysfonctionnements laryngés s'appuient sur la production de voyelles tenues, exercice qui permet d'obtenir une fiabilité de mesure mais qui s'éloigne des situations naturelles de communication.(Cisonni, 2008)

L'analyse acoustique de la voix et de la parole nécessite une bonne qualité du signal sonore, particulièrement dans l'étude des dysfonctionnements où il s'avère parfois nécessaire de mesurer de fines nuances qui ne doivent pas être entachées de bruit.(Rosso et al., 2008)

L'analyse la plus simple à effectuer à partir d'un enregistrement du signal audio est de représenter visuellement le son par les fréquences acoustiques qu'il contient et leur évolution au cours du temps.(Woo, 1969). L'oreille humaine est intégrative et elle ne permet pas toujours de distinguer avec précision les zones fréquentielles où l'énergie acoustique est renforcée ou atténuée.(Xu, 1994)

L'analyse acoustique conduit à extraire l'information verbale du signal vocal sous la forme de vecteur de paramètres acoustiques dont la dimension et la nature sont déterminants

pour accéder à de bonnes performances des systèmes. Les différents types de paramètres couramment cités dans la littérature sont : les coefficients de Cepstrum Mel-Frequency (MFCCs). (Babii & Ing, 1975).

L'extraction de la meilleure représentation paramétrique des signaux acoustiques est une tâche essentielle dans la conception de tout système de reconnaissance vocale. Il affecte remarquablement les performances de reconnaissance. Il existe plusieurs types de représentations paramétriques pour les signaux acoustiques. Parmi eux, les coefficients de Cepstrum Mel-Frequency (MFCC) La méthode la plus dominante et la plus largement utilisée pour extraire le spectre.

2.2 Paramètres MFCCs

2.2.1 Définition

Les coefficients de Cepstrum Mel-Frequency (MFCCs) sont l'une des techniques d'extraction de caractéristiques les plus populaires utilisées dans la reconnaissance vocale basée sur le domaine de fréquence en utilisant l'échelle Mel qui est basée sur l'échelle d'audition humaine, ce qui signifie qu'elle extrait des paramètres de la parole similaires à ceux utilisés par les humains pour entendre la parole, en même temps, réduisez toutes les autres informations.(Davis & Mermelstein, 1980)

Dans le domaine fréquentiel, les MFCCs sont la caractéristique la plus précise par rapport aux caractéristiques du domaine temporel. Ces caractéristiques MFCCs sont une représentation fidèle du signal fenêtré à courte portée dérivé de la transformée de Fourier rapide (FFT) de ce signal.(Combescure et al., 1995b)

Les coefficients MFCC sont un type de coefficients cepstraux très souvent utilisés en reconnaissance automatique de la parole.(F. Zheng et al., 2001)

2.2.2 Étapes de l'extraction des paramètres MFCCs

Les coefficients MFCC représentent les paramètres acoustiques que nous considérons pour représenter l'unité de son phonétique. En effet, ces coefficients permettent d'approximer le système auditif humain en regroupant les énergies des fréquences en bande de fréquences, et en dé corrélant le message transmis de l'appareil phonatoire humain.(Young et al., 2002)

La figure suivante illustre les différentes étapes du mécanisme d'extraction de ce paramètre acoustique du signal vocal.

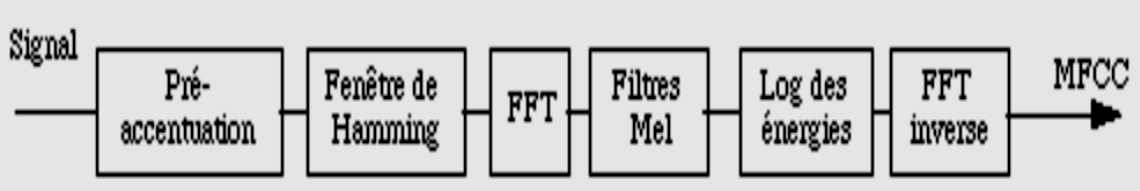


Figure II.1 : Étapes d'extraction des coefficients MFCC

- **Préaccentuation :**

La préaccentuation fait référence au filtrage qui accentue les fréquences plus élevées. Son but est d'équilibrer le spectre des sons vocaux qui ont un fort déclin dans la région des hautes fréquences. Pour les sons voisés, la source glottale a une pente d'environ -12 dB / octave. Cependant, lorsque l'énergie acoustique rayonne des lèvres, cela provoque une augmentation d'environ +6 dB / octave du spectre.(F. Zheng et al., 2001)

Dans Cette étape, le signal de parole est traité à travers un filtre qui accentue les fréquences plus élevées. Ce processus augmentera l'énergie du signal à une fréquence plus élevée.

$$Y[n] = X[n] - 0.95\ X[n-1] \qquad \textbf{(II.1)}$$

Où 0,95 signifie que 95% de tout échantillon devrait provenir de son échantillon précédent.

- **Fenêtrage :**

Le signal vocal est d'abord divisé en tranches de temps constituées de N échantillons dont la longueur est comprise entre 20 et 40 msec. Les trames adjacentes sont séparées par M (M <N) valeurs standard utilisées sont M = 100 et N = 256 échantillons. (F. Zheng et al., 2001)

- **Fenêtre de Hamming :**

Chaque fenêtre temporelle est ensuite fenêtrée avec une fenêtre de Hamming pour exclure les discontinuités sur les bords (Nitisha & Bansal, 2012). L'équation de la fenêtre de Hamming est donnée comme suit :

$$W(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right),\ 0 \le n \le N-1 \qquad \textbf{(II.2)}$$
$$= 0, \text{otherwise}$$

Où:
-la fenêtre est définie comme W (Pols)
-n = est l'échantillon actuel.
-N = est le nombre d'échantillons dans chaque trame
-Y [n] = signal de sortie
-X (n) = signal d'entrée
-W (n) = Fenêtre de Hamming
Le signal de fenêtrage est alors: $Y(n) = X(n) * W(n)$ **(II.3)**

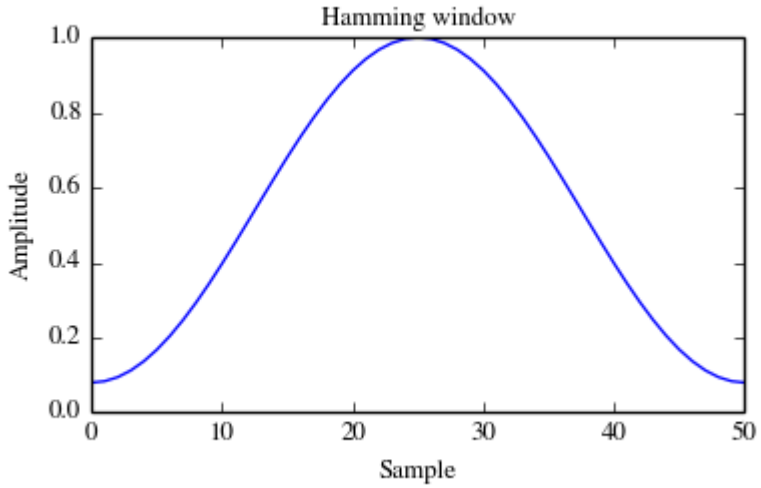


Figure II.2 : Hamming Window.

➢ **Transformation de Fourier rapide (Fast Fourier Transformation (FFT)) :**

La transformation de Fourier rapide (acronyme anglais : FFT ou Fast Fourier Transform) est un algorithme de calcul de la transformée de Fourier discrète (TFD). Ainsi, pour le temps de calcul de l'algorithme rapide peut être 100 fois plus petit que le calcul utilisant la formule de définition de la TFD. Cet algorithme est couramment utilisé en traitement numérique du signal pour transformer des données du domaine temporel au domaine fréquentiel. (Ittichaichareon, 2012).

➢ **Spectre DFT (DFT Spectrum):**

Chaque trame fenêtrée est convertie en spectre de magnitude en appliquant la DFT.

$$X(k) = \sum x(n)\, e^{\frac{-2\pi nkj}{N}} \; ; 0 \leq k \leq N-1 \qquad \textbf{(II.3)}$$

Où N est le nombre de points utilisés pour calculer la TFD.

➢ **Traitement de la banque de filtres Mel**

La banque de filtres logarithmiques Mel-Scaled est appliquée à la trame transformée de Fourier. La figure II.3 représente la forme générale du banc de filtres, un ensemble de filtres triangulaires.

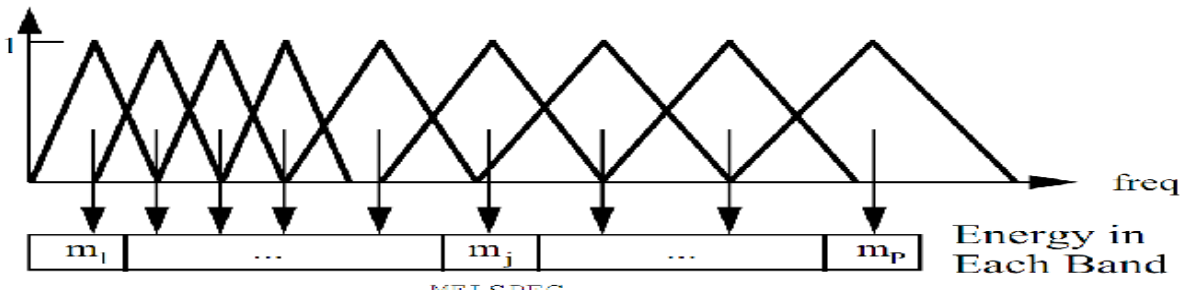


Figure II.3 : Banque de filtres Mel-Scale

La relation entre la fréquence de la parole et l'échelle de Mel peut être représentée comme suit:

$$M(f) = 2595\log10\,(1 + f\,(Hz)/700) \quad \textbf{(II.4)}$$

Pour revenir de Mels à la fréquence:

$$M(m) - 1 = 700\,(\exp\,(m/2595) - 1) \quad \textbf{(II.5)}$$

- **Spectre Mel :**

Le spectre Mel est calculé en faisant passer le signal transformé de Fourier à travers un ensemble de filtres passe-bande appelés banc de filtres Mel. Un Mel est une unité de mesure basée sur la fréquence perçue par les oreilles humaines. Il ne correspond pas linéairement à la fréquence physique du son, car le système auditif humain ne perçoit apparemment pas la hauteur linéairement.

- **Transformation discrète en cosinus (DCT):**

Le DCT appliqué aux coefficients de fréquence Mel transformés produit un ensemble de coefficients cepstraux. Avant de calculer le DCT, le spectre Mel est généralement représenté sur une échelle logarithmique. Il en résulte un signal dans le domaine cepstral avec un pic de fréquence que correspondant à la hauteur du signal et un certain nombre de formants représentant des pics de fréquence faible.

$$\mathbf{C_i = \sqrt{2/N}\ \sum m_j\,.\,cos\,[\pi_i / N(j - 0{,}5)]} \quad \textbf{(II.6)}$$

N est le nombre de canaux de banque de filtres.

Cet ensemble de coefficients est appelé un vecteur acoustique. Par conséquent, chaque énoncé d'entrée est transformé en une séquence de vecteurs acoustiques. (S. Young, 1997).

- **Fonctionnalités MFCC dynamiques:**

Les coefficients cepstraux sont généralement appelés entités statiques, car ils ne contiennent que des informations provenant d'une trame donnée. Les informations supplémentaires sur la dynamique temporelle du signal sont obtenues en calculant les première et deuxième dérivées des coefficients cepstraux.

La dérivée de premier ordre est appelée coefficients delta et la dérivée de second ordre est appelée coefficients delta – delta. Les coefficients delta indiquent le débit de la parole et les coefficients delta fournissent des informations similaires à l'accélération de la parole.

2.3 Discrimination automatique

2.3.1 Deep Learning

2.3.1.1 Définition (généralité)

Le Deep Learning ou apprentissage profond est un type d'intelligence artificielle dérivé de la machine Learning (apprentissage automatique) où la machine est capable d'apprendre par elle-même, contrairement à la programmation où elle se contente d'exécuter à la lettre des règles prédéterminées. « La technologie du Deep Learning apprend à représenter le monde. C'est-à-dire comment la machine va représenter la parole ou l'image par exemple ».(Deng & Yu, 2013)

Le cerveau humain est «profond» que toute action est le résultat de longues connexions synaptiques avec de nombreuses couches de traitement. Le Deep Learning combine une classe d'algorithmes d'apprentissage correspondant à ces structures profondes. Il est souvent utilisé pour l'apprentissage, c'est-à-dire l'apprentissage simultané des propriétés bénéfiques des données et de la meilleure façon de les utiliser. [8]

2.3.1.2 Avènement du Deep Learning

Ces dernières années, la science a progressé dans le domaine des performances de calcul des ordinateurs et avec elle le concept d'apprentissage profond a évolué. Ce sont des réseaux de neurones avec de nombreux neurones situés entre les couches d'entrée, qui acceptent les données à traiter, et les couches de sortie, qui visent à présenter le résultat d'un calcul.(Arnold et al., 2010) À la grande surprise des spécialistes, l'ajout de ces couches de neurones a eu un effet très bénéfique sur la qualité des résultats obtenus. C'est ce qui a permis à l'intelligence artificielle de faire un retour ces dernières années. La plupart des acteurs du domaine ne jurent plus que par le DL.(van Gogh, 2014)

2.3.1.3 Origine de Deep Learning

Dans les années 1950, le mathématicien britannique Alan Turing imagine une machine capable d'apprendre, une « Learning Machine ». Au cours des décennies suivantes, différentes techniques de Machine Learning ont été développées pour créer des algorithmes capables d'apprendre et de s'améliorer de manière autonome.[9]

2.3.2 De Machine Learning au Deep Learning

L'apprentissage automatique (ML) offre une variété de technologies et de modèles parmi lesquels nous pouvons choisir, en fonction de l'application, du volume de données que nous traitons et du type de problème que nous voulons résoudre, par exemple: à mesure que les données augmentent, les paramètres s'affinent, et le biais du modèle est réduit. D'une manière plus efficace. L'apprentissage en profondeur (DL) est un sous-domaine plus avancé de l'apprentissage automatique. Un apprentissage en profondeur réussi nécessite une très grande quantité de données pour entraîner le modèle, en plus d'une formidable capacité de

[8] Article deep Learning sur la plateforme de data analytics post publier le 5 février 2014
[9] Article histoire deep learning publie sur la plateforme naturel science publier le 19 mars 2002

calcul pour traiter les données en un minimum de temps. Pour choisir entre l'apprentissage automatique et l'apprentissage en profondeur, nous devons savoir autant que possible si nous disposons d'un GPU hautes performances et de nombreuses données étiquetées. Si nous n'avions aucune de ces choses, il pourrait être judicieux d'utiliser l'apprentissage automatique plutôt que l'apprentissage profond. L'apprentissage en profondeur est généralement plus complexe, nous avons donc besoin de plus de données pour obtenir des résultats fiables. Cela signifie d'avoir un GPU haute performance qui permet au modèle de prendre moins de temps pour analyser toutes ces données.

2.3.3 Fonctionnement de Deep Learning

Dans le cerveau humain, chaque neurone reçoit environ 100 000 signaux électriques d'autres neurones. Chaque neurone actif peut produire un effet excitateur ou inhibiteur sur celui avec lequel il est lié. A l'intérieur d'un réseau artificiel, le principe est le même. Les signaux voyagent entre les cellules nerveuses. Le réseau neuronal détermine un poids spécifique pour différents neurones. Et les neurones qui reçoivent plus de charge ont un effet plus important sur les neurones voisins. La dernière couche de neurones émet une réponse à ces signaux.[10]

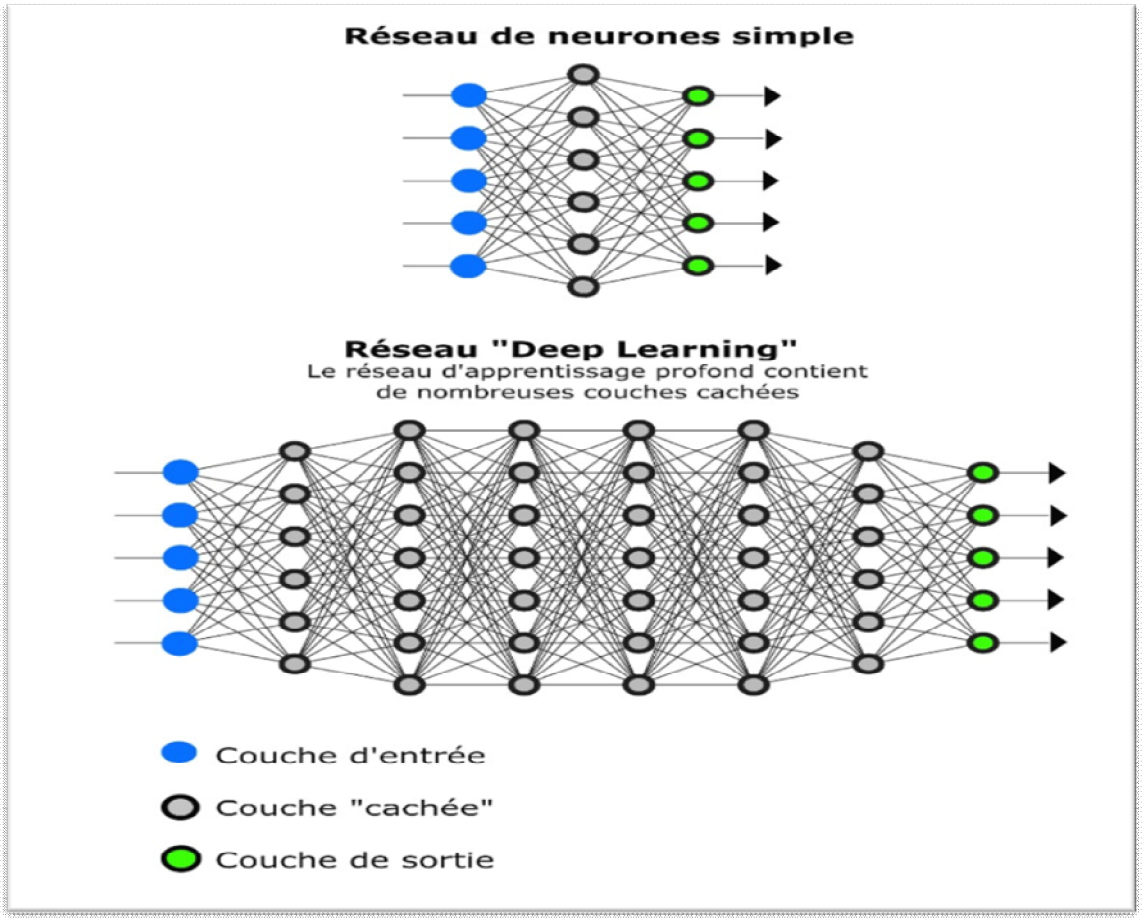


Figure II.4 : Réseau de neurone ET Réseau Deep Learning (Arnold et al., 2010)

[10] Article deep Learning publié sur la plateforme de lebiG data

DL a de nombreuses utilités. C'est cette technologie qui est utilisée pour la reconnaissance faciale de la reconnaissance d'images par exemple. Depuis quelques années, avec l'apparition des réseaux de neurones convolutifs, DL est au cœur de la vision par ordinateur et de la vision robotique. Comme l'explique le professeur Peter Corke, étant donné que les réseaux de neurones artificiels imitent le fonctionnement du cerveau humain, les possibilités offertes par cette technologie augmenteront à mesure que nous découvrons les secrets de notre propre organe. (Arnold et al., 2010)

L'utilisation de DL a considérablement augmenté au cours des dernières années avec l'essor des GPU, des Big Data, des fournisseurs de Cloud tels qu'Amazon Web Services (AWS) et Google Cloud, et des frameworks tels que Torch, TensorFlow, Caffe et PyTorch. En plus de cela, les grandes entreprises partagent des algorithmes formés sur d'énormes ensembles de données, aidant ainsi les startups à construire des systèmes de pointe sur plusieurs cas d'utilisation avec peu d'effort.(Arnold et al., 2010)

2.3.4 Exemple de Deep Learning

Il y a nombreux des problèmes réels qui sont résoudra par DL. Ce dernier aide les machines de faire détection, classification et aussi identifier des objets et traduire les paroles…etc. voilà quelques exemples de ces problèmes :

2.3.4.1Conduite automatisé

L'automatisation est devenue une composante importante de nombreuses industries au XXIe siècle, qui comprennent les domaines de la fabrication, des communications et des transports. L'automatisation a fourni de nombreux retours et résultats, notamment une sécurité et une productivité améliorée et une réduction des coûts, par exemple l'application de DL pour détecter les piétons et aider à réduire le taux d'accidents. Les chefs de file de l'industrie se sont consacrés à la mise en œuvre de technologies autonomes dans le transport de production, telles que les véhicules «sans conducteur» ou les véhicules entièrement autonomes, et la détection automatique d'éléments tels que les panneaux d'arrêt, les feux de signalisation, etc. La technologie clé qui a le potentiel de conduire le futur développement technologique de ces véhicules est l'apprentissage en profondeur.(Bock & Garnsey, 2008)

2.3.4.2Défense et aérospatial Deep Learning

Certain algorithmes sont utilisés pour la détection et l'identification des différents objets à partir de satellites ces derniers peuvent localisé les intrusions et les zones sûres et dangereuses.(Bock & Garnsey, 2008)

2.3.4.3Recherche médicale

Dans le domaine de la médecine, les techniques d'apprentissage en profondeur ont été largement utilisées et en profondeur, ce qui a donné des résultats impressionnants et sont encourageants, en particulier pour les grandes bases de données. Un algorithme d'apprentissage en profondeur via un réseau neuronal de contournement a été appliqué dans des images médicales à l'aide d'images échographiques sur un groupe de nerfs connu sous le nom de plexus brachial. Une technique d'apprentissage en profondeur est également

appliquée pour classer les différents stades de la rétinopathie diabétique à l'aide de l'imagerie de fond couleur de la rétine, et il y a autre utilisation comme classification des maladies …etc.(Bock & Garnsey, 2008)

2.3.4.4Automatisation industrielle

Les capacités d'apprentissage en profondeur ont atteint la découverte des relations sémantiques dans le langage écrit naturel, car elles reposent sur l'intervention humaine pour fournir les paramètres nécessaires. Une représentation précise et satisfaisante des documents est la clé pour soutenir les modèles informatiques pour arriver à la signification de base dans la langue écrite. La classification de texte est le sac de mots sur lequel vous comptez pour mettre en place une approche représentationnelle, dans laquelle les documents sont représentés par un vecteur de mots qui apparaissent dans chaque document. Bien que le sac de mots soit facile à créer, la stratégie principale permet la présentation du vecteur résultant qui est très grand et clairsemé. Cette difficulté et la nécessité d'assurer une compréhension sémantique des documents textuels sont les principaux défis de la classification des textes. L'apprentissage en profondeur a permis d'obtenir de nombreuses possibilités et méthodes pour représenter des mots et des documents, parmi lesquels figurent Word2Vec, Doc2Vec, GloVe, LSTM et CNN. (Bock & Garnsey, 2008)

2.3.4.5Traduction

Le développement de capacités et de méthodes d'apprentissage en profondeur permet à l'utilisation de Google Translate d'être prise en charge par un réseau de mémoire à court terme de 8 couches avec une communication et un intérêt résiduels.[11]

2.3.4.6Reconnaissance vocale au niveau quasi humain

Au cours des dernières années, le réseau de neurones MLP traditionnel a été utilisé pour la reconnaissance vocale, mais ses performances sont inférieures à celles des systèmes HMM avancés avec des probabilités de surveillance approximatives utilisant GMM. Avec le développement des techniques de l'apprentissage en profonde, il a été intégré avec succès dans le réseau de neurones MLP pour des tâches de vocabulaire téléphonique et de reconnaissance vocale étendues. Ils sont intégré aussi également la forte capacité de formation discriminante des DNN avec la capacité de modélisation en série des HMM pour obtenir le meilleur(Bock & Garnsey, 2008)

2.3.5 Les Algorithmes de Deep Learning

Le Deep Learning propose de se placer directement au niveau des données sans déformation ou agrégation préalable. Puis, au moyen d'un nombre extrêmement grand de paramètres qui s'auto-ajustent au fil de l'apprentissage, le réseau va apprendre de lui-même les liens implicites qui existent dans les données. Il y a différents algorithmes utilisés en apprentissage profond pour différents cas d'usage, Comme il est indiqué dans l'image suivante

[11](https://robots.net/tech-reviews/deep-learning-applications/#:~:text=Applications%20of%20deep%20learning%20have,board%20games%20programs%20and%20material)

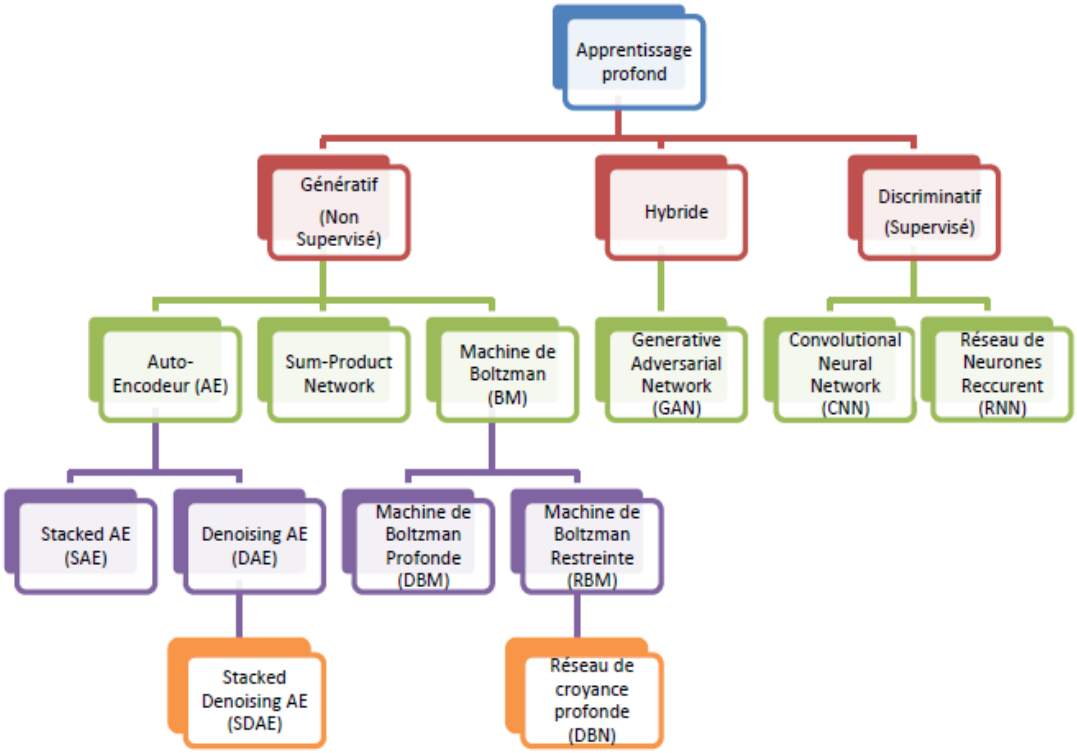


Figure II.5 : Classification des méthodes de L'apprentissage profond

Nous commençons par définir simplement le modèle au cœur de l'apprentissage profond : le "réseau de neurones".

2.3.5.1 Réseau de neurones

Les réseaux de neurones n'ont que très peu à voir avec le système neuronal et le cerveau. L'analogie entre un neurone et un réseau de neurones à une couche est essentiellement graphique, dans la mesure où il y a un flux d'information d'un bout à l'autre du réseau.

Il existe un grand nombre de modèles d'architecture profonde. La plupart d'entre eux sont dérivés de certaines architectures parentales originales. Il n'est pas toujours possible de comparer les performances de toutes les architectures, car toutes ne sont pas évaluées sur les mêmes ensembles de données. DL est un domaine en pleine croissance et de nouvelles architectures, variantes ou algorithmes apparaissent chaque semaine. Les types de réseaux de neurones profonds les plus populaires sont le réseau neuronal convolutif (CNN), nous présentons comme suite :

2.3.5.2 Réseau Neuronal Convolutif (CNN)

Les CNN sont les plus couramment appliqués à la segmentation et à la classification des images. Les CNN ont été introduits pour la première fois en 1989, mais ont suscité un grand intérêt après que des CNN profonds ont obtenu des résultats spectaculaires dans la compétition Image Net en 2012. Appliqués sur un ensemble de données d'environ un million d'images comprenant 1000 classes différentes, les CNN ont presque réduit de moitié les taux

d'erreur des meilleures approches informatiques auparavant.(Bar et al., 2015). Les CNNs sont appliqué sur différent domaine telle que la reconnaissance et la classification d'images, la détection d'objets, la segmentation et d'autres nombreuses tâches dans le domaine de la vision par ordinateur, et dans le domaine du traitement du langage naturel (NLP)…etc.

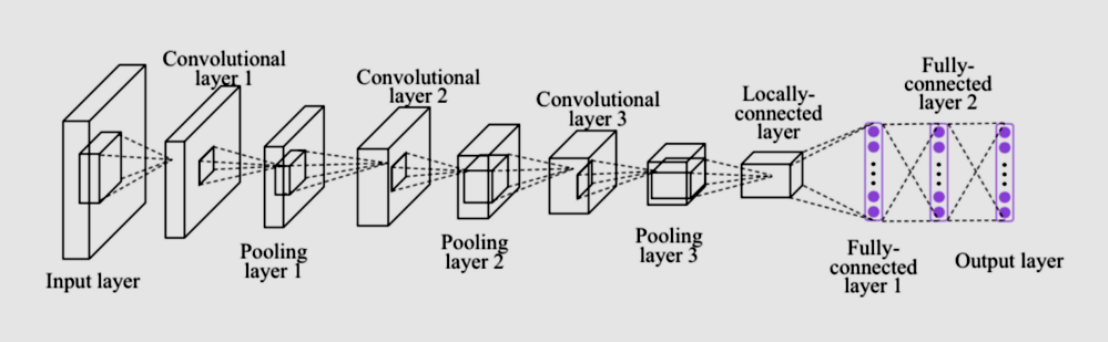


Figure II.6 : Schéma d'un réseau de neurones convolutionnel [12]

Comme les autres réseaux de neurones, un CNN est composé d'une couche d'entrée, d'une couche de sortie et de nombreuses couches cachées comme suit :

➢ Couche convolutionnelle

Trois hyper paramètres permettent de dimensionner le volume de la couche de convolution : la profondeur, le pas et la marge.

- **Profondeur de la couche** :

Nombre de noyaux de convolution (ou nombre de neurones associés à un même champ récepteur).

- **Le pas** :

Contrôle le chevauchement des champs récepteurs. Plus le pas est petit, plus les champs récepteurs se chevauchent et plus le volume de sortie sera grand.

- **La marge (à 0) ou (zero padding)** :

Parfois, il est commode de mettre des zéros à la frontière du volume d'entrée. La taille de ce 'zero-padding' est le troisième hyper paramètre. Cette marge permet de contrôler la dimension spatiale du volume de sortie. En particulier, il est parfois souhaitable de conserver la même surface que celle du volume d'entrée.

Des filtres sont appliqués à chaque image utilisée pour l'apprentissage à différentes résolutions, et la sortie de chaque image convolutée est utilisée comme entrée de la couche suivante. Dans la terminologie du réseau convolutif, le premier argument de la convolution est souvent appelé l'entrée (input) et le second argument comme noyau (kernel). La sortie est parfois appelée la carte des caractéristiques (feature map).

[12] https://www.groundai.com/project/a-survey-on-deep-learning-based-brain-computer-interface/

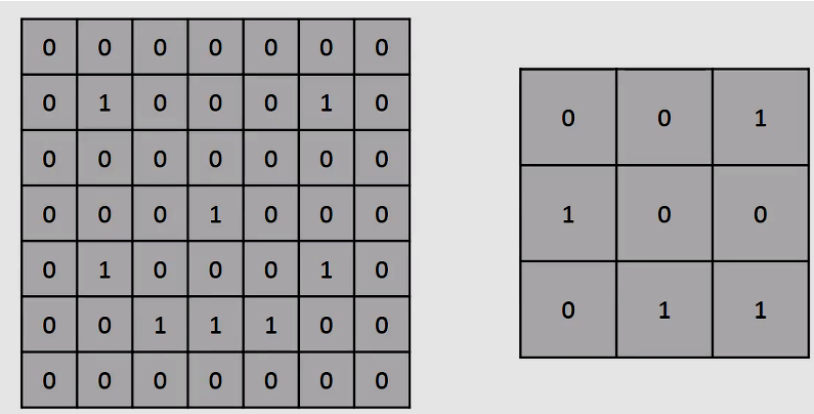


Figure II.7 : Matrices de l'image input de pixels et filtre [13]

La figure. II.8 représente l'opération de convolution de l'image 4 x 4 et de la matrice 2 x 2. La matrice de sortie est appelée fonction convolutée. Les valeurs différentes de la matrice de filtre produiront différentes cartes de caractéristiques pour la même image d'entrée. « Plus nous avons de filtres, plus nous obtenons de fonctionnalités et plus notre réseau reconnaît mieux les motifs dans les images ».

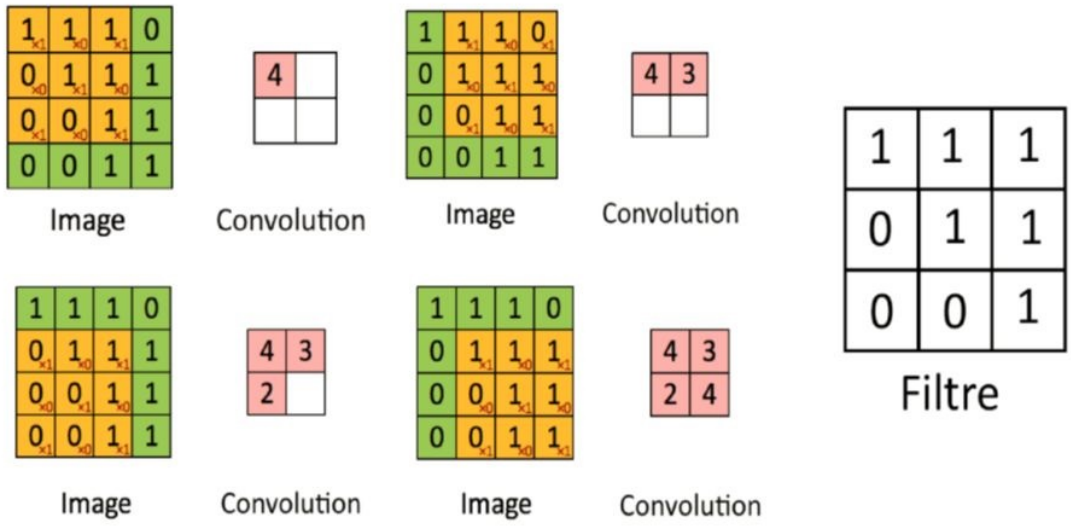


Figure II.8 : la fonction convolutée.[5]

[13] Article Under standing of Convolutional Neural Network (CNN) — Deep Learning publie sur la plateforme de medium le 04 mars 2017

➢ Fonction d'activation

Applique une fonction non linéaire telle que (Unité linéaire rectifiée**(ReLU))** et elle est appliquée à chaque élément des cartes d'entités (aux sorties) de la couche convolutionnelle et remplace toutes les valeurs de pixels négatives par zéro comme dans la figure II.9 suivante :

$$\text{ReLU} == > R(x) = \max(0, z) \qquad \textbf{(II.7)}$$

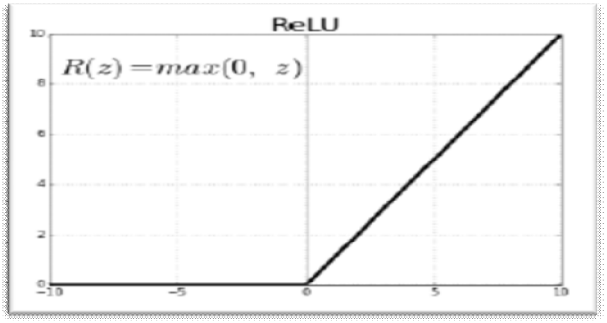


Figure II.9 : la fonction ReLU [14]

Il existe d'autres fonctions que ReLU telles que tanh ou sigmoid.

➢ Couche de mise en commun

Applique une opération de sous-échantillonnage qui réduit la taille des cartes d'entités et les résultats de la couche de convolution avec une opération Max () ou Sum () ou Moyenne (). Le regroupement offre deux fonctionnalités différentes : l'une réduit la taille des données à traiter et l'autre force l'algorithme à ne pas se concentrer sur les petits changements de position d'une image.

Exemple : si nous appliquons Max Pooling, nous définissons une fenêtre spécifique et prenons le plus grand élément de la carte d'entités ReLU de cette fenêtre comme la figure II.10.

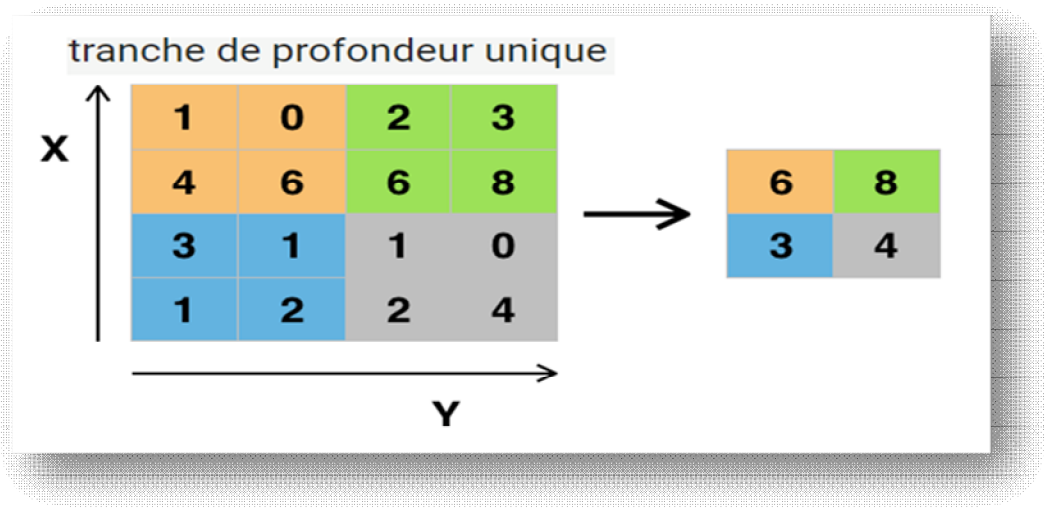


Figure II.10 : Exemplaire de Max Pooling [5]

[14] https://mc.ai/activation-functions-relu-softmax/

➢ Couche entièrement connectée (classification)

Chaque neurone de la couche précédente est connecté à chaque neurone de la couche suivante, cette dernière couche produisant des sorties de réseau neuronal. Soft max est la fonction d'activation dans la couche de sortie de la couche entièrement connectée.

La formule calcule l'exponentielle de la valeur d'entrée donnée et la somme des valeurs exponentielles de toutes les valeurs des entrées. Le rapport de l'exponentielle de la valeur d'entrée et de la somme des valeurs exponentielles est alors la sortie de la fonction soft max.

$$f\ (x_i) = \frac{\text{Exp}\ (x_i)}{\sum_{j=0} \text{Exp}\ (x_j)} \qquad \textbf{(II.8)}$$

Nous ajoutons la normalisation par lots pour augmenter la stabilité et l'abandon pour éviter le sur-ajustement. Nous expliquons comment fonctionne l'architecture :

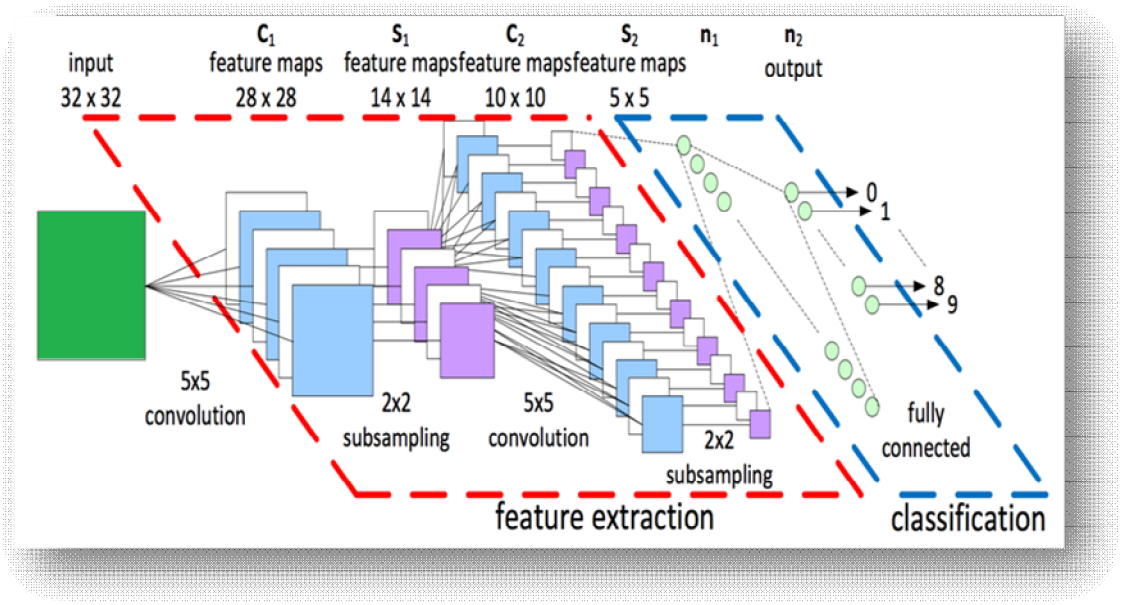


Figure II.11 : L'architecture de CNN

2.3.6 Réseaux de neurones convolutifs pour la classification d'images

La classification d'images a été l'un des sujets les plus importants dans le domaine de la vision par ordinateur, visant à donner aux machines la capacité de reconnaître des modèles et de s'adapter à différents environnements d'image. La reconnaissance faciale, la détection des véhicules, le diagnostic médical et la reconnaissance des chiffres sont des exemples typiques.

Les réseaux de neurones convolutifs ont été largement utilisés dans ce domaine. Voici quelques exemples exploitant les CNNs pour la classification d'images:

• Le travail d'Alex Krizhevskyet al. (J. Sun et al., 2017) visant la création d'un «grand réseau de neurones convolutifs profonds» pour la classification d'images. Le modèle résultant est le premier modèle performant sur l'ensemble de données ImageNet.

• Le Cun Yann, et al. (Cid et al., 2017) passent en revue diverses méthodes appliquées à la reconnaissance de caractères manuscrits en utilisant un CNN. Ce travail a exploité la base de données MNIST contenant un grand ensemble de chiffres manuscrits.

• S. Lawrence et al. (Guerfi, 2011) explorent l'idée d'utiliser un ConvNet pour la reconnaissance faciale afin d'extraire des fonctionnalités successivement plus grandes dans un ensemble hiérarchique de couches.

• Chuan Zhou et al (R. Sun et al., 2017) ont appliqué les CNNs à l'imagerie médicale en classant les régions d'intérêt sur les mammographies en masse ou en tissu normal.

2.3.7 Apprentissage Par Transfert

Le transfert d'apprentissage consiste à transférer les paramètres du modèle entraîné vers le nouveau modèle pour aider le nouveau modèle à s'entraîner. Étant donné que certaines données ou tâches sont pertinentes, l'efficacité d'apprentissage du nouveau modèle peut être accélérée et optimisée en partageant les paramètres du modèle entraîné avec le nouveau modèle grâce à l'apprentissage par transfert, au lieu de partir de zéro comme le font la plupart des réseaux.

Ces dernières années, en raison de la popularité de l'apprentissage profond, la demande de données augmente également, de sorte que l'apprentissage par transfert a été développé davantage. Afin de former un classificateur de haute qualité à partir des données de différentes distributions

Dai et coll. (Dai et al., 2007) ont proposé un algorithme d'apprentissage de transfert TrAdaBoost basé sur des instances, qui utilise les données d'apprentissage pour sélectionner les parties utiles à partir des différentes données de distribution, afin de réduire le taux d'erreur de classification. Afin de résoudre le problème de l'adaptation de domaine dans l'apprentissage par transfert.

Cao et coll. (Cao et al., 2018.) Ont proposé un apprentissage par transfert partiel basé sur des réseaux contradictoires sélectifs (SAN). Il n'a transféré que les échantillons partiels liés au domaine source et au domaine cible, puis a traité les problèmes de transfert partiel via SAN. Étant donné que le réseau contradictoire peut bien apprendre les fonctionnalités invariantes de domaine, il peut jouer un grand rôle dans l'apprentissage par transfert.

Busto et Gall (Busto & Gall, 2017) ont proposé une adaptation de domaine ouvert. Ils utilisent la relation entre le domaine source et le domaine cible pour étiqueter les échantillons du domaine cible et convertir le domaine source dans le même espace que le domaine cible, et enfin classer les échantillons du domaine cible.

2.3.7.1 Modèle VGG16

Le réseau VGG a été présenté par les chercheurs du Visual Graphics Group à Oxford (d'où le nom VGG) en 2014. Et présente par Simonyan et Zisserman dans leur article de 2015, (Simonyan & Zisserman, 2015). Ce réseau est spécialement caractérisé par sa forme pyramidale, où les couches inférieures les plus proches de l'image sont larges, tandis que les couches supérieures sont profondes.

Comme montré dans la figure II.12, VGG16 contient des couches convolutives ultérieures suivies par des couches de regroupement. Les couches de regroupement sont responsables de rendre les couches plus étroites.

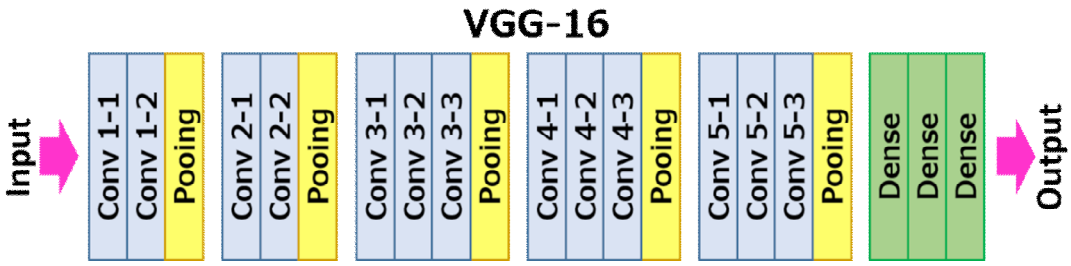


Figure II.12 : L'architecteur de modèle VGG16

2.3.7.2 Modèle Xception

Xception est une architecture de réseau neuronal à convolution profonde qui implique des convolutions séparables en profondeur. Il a été développé par des chercheurs de Google. Xception est un réseau neuronal convolutif de 71 couches de profondeur. Définie comme un modèle pré-entraînée du réseau formée sur plus d'un million d'images de la base de données ImageNet. En conséquence, le réseau a appris de riches représentations de caractéristiques pour une large gamme d'images.(Chollet, 2017). Le réseau a une taille d'entrée d'image de 299 x 299 x 3.

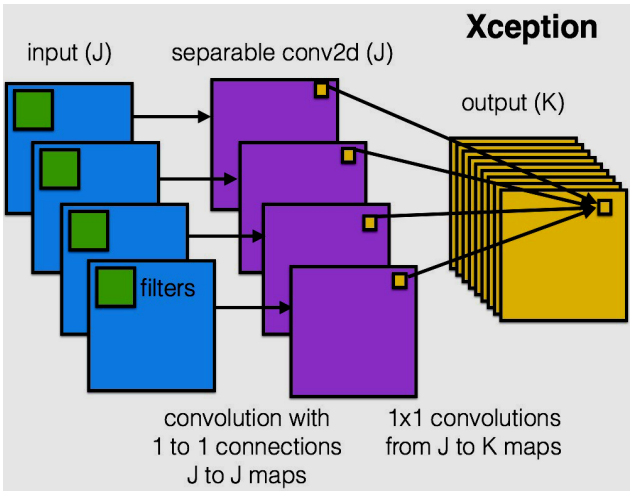


Figure II.13 : L'architecteur de modèle Xception

2.3.7.3 Modèle Inception V3

La microarchitecture «Inception» a été introduite pour la première fois par Szegedy et al, dans leur article de 2015 (Szegedy et al., 2015). Ce module est un «extracteur de caractéristiques à plusieurs niveaux» en calculant des convolutions 1×1, 3×3 et 5×5 dans le même module du réseau, la sortie de ces filtres est alors empilée le long la dimension du canal et avant d'être introduit dans la couche suivante du réseau. L'incarnation originale de cette architecture s'appelait GoogLeNet, mais les manifestations ultérieures ont été appelées Inception vN où N se réfère au numéro de version publié par Google.

L'architecture Inception V3 incluse dans le noyau Keras provient de la publication ultérieure de Szegedy et al qui propose des mises à jour du module de démarrage pour améliorer encore la précision de classification d'ImageNet. (Szegedy et al., 2016)

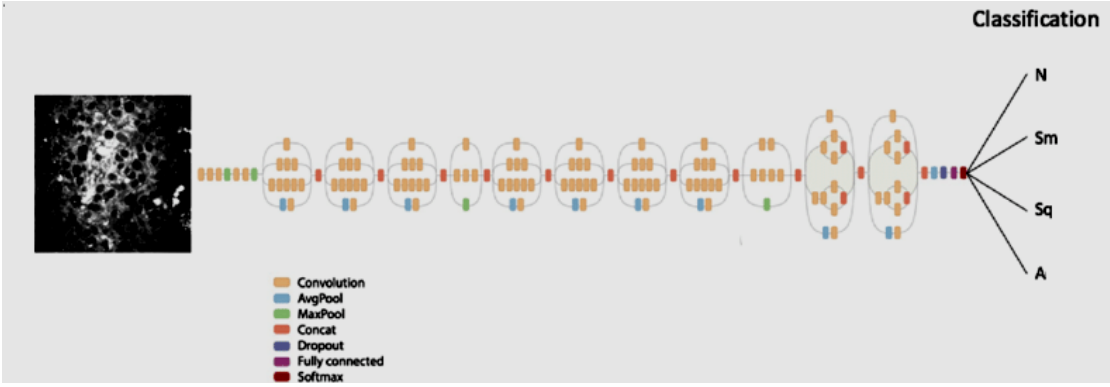


Figure II.14 : L'architecteur de modèle Inception V3

2.3.7.4 Modèle ResNet

ResNet est l'abréviation du nom composé RESidiual Network (Réseaux résiduels). ResNet contrairement aux architectures de réseau séquentielles traditionnelles telles qu'AlexNet, OverFeat et VGG, ResNet est plutôt une forme d '«architecture exotique» Présenté pour la première fois par He et al, dans leur article de 2016. (He et al., 2016)

ResNet est l'une des architectures qui définissent véritablement la profondeur d'une architecture d'apprentissage en profondeur. Les ResNets sont constitués de plusieurs modules résiduels subséquents, qui constituent le bloc de construction de base de l'architecture ResNet. Ces modules résiduels sont empilés les uns sur les autres pour former un réseau complet. Dans la figure II.16 la représentation de l'architecture ResNet

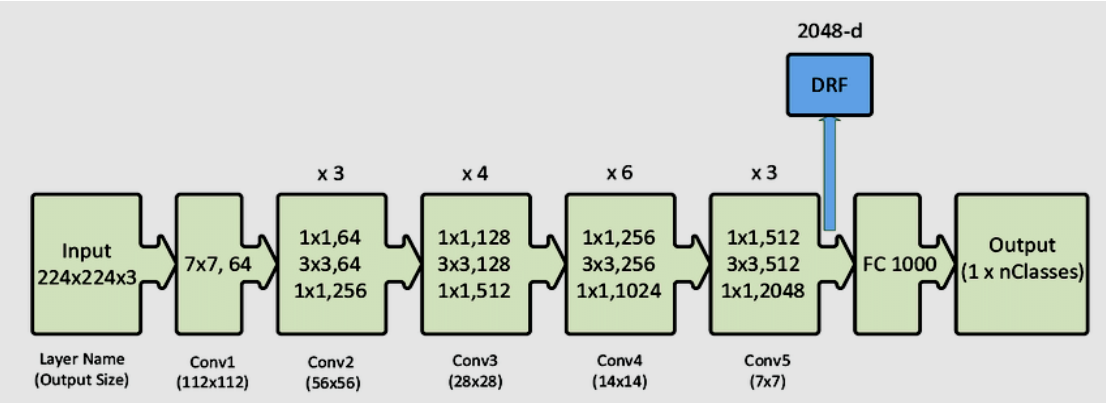


Figure II.15 : L'architecteur de modèle ResNet

3. Mesures d'évaluation des modèles

En termes d'évaluation des modèles, plusieurs mesures ont été définies dans le but de valoriser la précision et la robustesse des différents algorithmes d'apprentissage automatique ou profond. En général, positif = identifié et négatif = rejeté.(Tan, 2017) Donc :

- **Vrai positif :** ***TP*** = correctement identifié
- **Faux positif :** ***FP*** = identifié de manière incorrecte
- **Vrai négatif :** ***TN*** = correctement rejeté
- **Faux négatif :** ***FN*** = rejeté de manière incorrecte
- **Rappel ou Sensibilité ou TPR (Taux Vrai Positif) :** Nombre d'éléments correctement identifiés comme positifs par rapport au total des vrais positifs. Cette mesure est donnée par la formule :

$$Recall = TPR = TP / (TP + FN) \quad \textbf{(II.9)}$$

- **Spécificité ou TNR (Taux Vrai Négatif) :** nombre d'éléments correctement identifiés comme négatifs par rapport au total des négatifs.

$$TNR = TN / (TN + FP) \quad \textbf{(II.10)}$$

- **Précision :** Nombre d'éléments correctement identifiés comme positifs sur le total des éléments identifiés comme positifs.

$$Précision = TP / (TP + FP) \quad \textbf{(II.11)}$$

- **Taux de faux positifs ou erreur de type I :** nombre d'éléments incorrectement identifiés comme positifs par rapport au total des véritables négatifs.

$$FPR = FP / (FP + TN) \quad \textbf{(II.12)}$$

- **Faux taux négatif ou erreur de type II :** nombre d'éléments identifiés à tort comme négatifs par rapport au total des vrais positifs.

$$FNR = FN / (FN + TP) \quad \textbf{(II.13)}$$

➢ **Matrice de confusion :** également connue sous le nom de matrice d'erreur, est une disposition de tableau spécifique permettant de visualiser les performances d'un algorithme, généralement un algorithme d'apprentissage supervisé (dans le cas d'un apprentissage non supervisé, il s'agit généralement d'une matrice correspondante). (Leclercq, 2013)

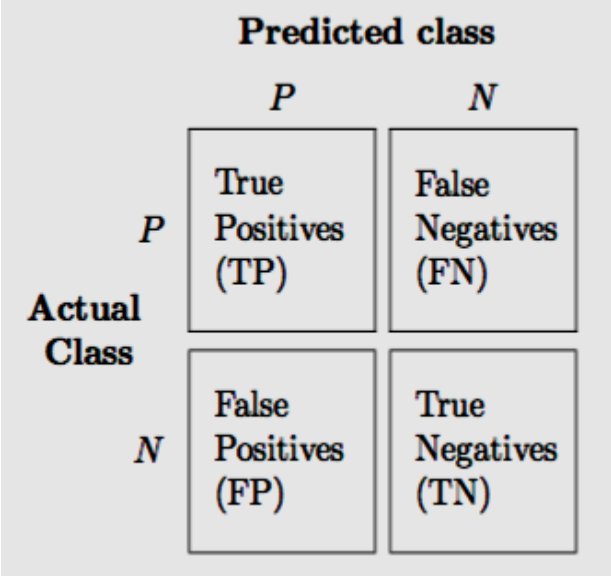


Figure II.16 : matrice de confusion

➢ **F1 Score:** C'est une moyenne harmonique de précision et de rappel[15] donnée par :

$$F1\ Score = 2 * Precision * Recall / (Precision + Recall) \quad \textbf{(II.14)}$$

➢ **Exactitude (Accuracy) :** pourcentage du nombre total d'éléments classés correctement.[9]

$$Accuracy = (TP + TN) / (N + P) \quad \textbf{(II.15)}$$

➢ **ROC-AUC Score**

Est un graphe qui montre la performance d'un modèle pour tous les seuils, Il s'agit du taux des vrais positifs en fonction de celui des faux positifs. La courbe ROC montre ces taux pour différents seuils et Plus cette courbe est éloignée de la première bissectrice, plus le modèle à une forte capacité à discriminer. (Tarayoun, 2018)

L'indicateur synthétique associé à la courbe ROC est la surface située sous la courbe, c'est l'AUC (Aera Under the Curve). Un modèle de classification est performant si l'AUC est proche de 1. À l'inverse, un modèle de classification n'est pas discriminant si l'AUC est proche de 0,5.[16]

[15] (https://www.memoireonline.com/01/12/5185/m_Mesures-et-determinants-de-laconfiance-des-menages-sur-la-situation-economique-au-Cameroun-cas28.html)

[16](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjtriXjIPjAhVwD2MBHaIMDiAQFjAAegQIABAB&url=https%3A%2F%2Fopenclassrooms.com%2Ffr%2Fcourses%2F4470531-classez-et-segmentez-des-donnees-visuelles%2F5072281-utilisez-ces-features-pour-classifier-desimages&usg=AOvVaw3ofidMifp1cojMiUpOKiau)

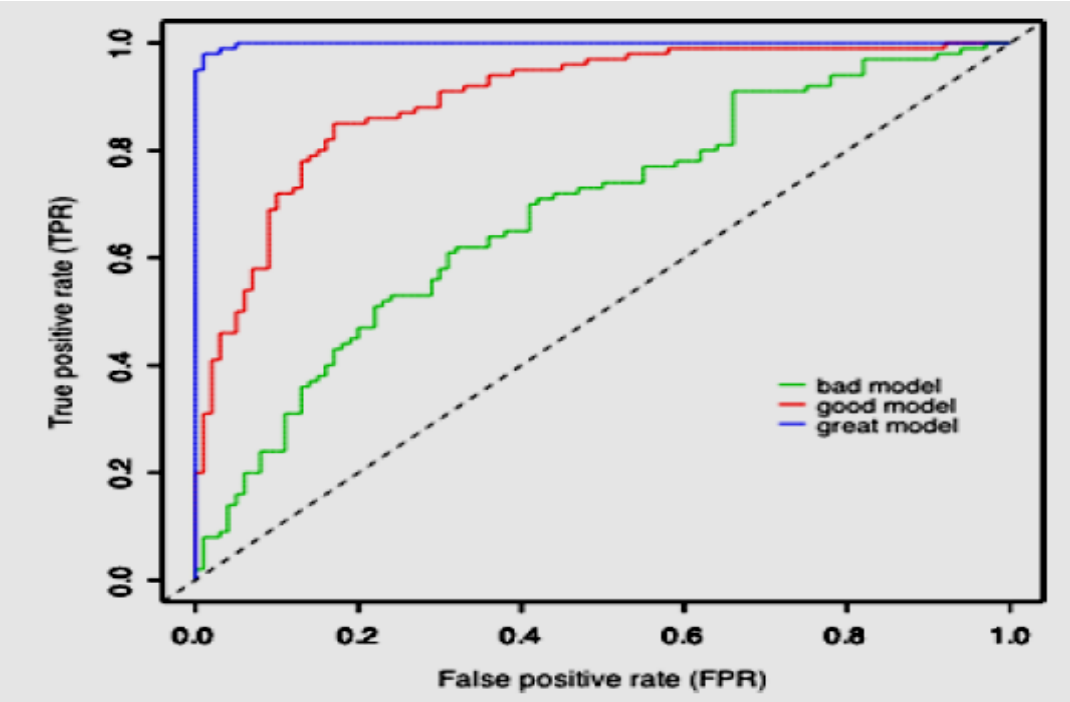


Figure II.17 : Interprétation graphique de la mesure ROC-AUC.

Certaines caractéristiques importantes de ROC-AUC sont:

- La valeur peut aller de ***0*** à ***1***. Toutefois, le score de AUC d'un classifieur aléatoire pour des données équilibrées est de ***0.5***.
- Le score ROC-AUC est indépendant du seuil défini pour la classification car il ne prend en compte que le rang de chaque prédiction et non sa valeur absolue. Il n'en va pas de même pour le F1 score qui nécessite une valeur seuil en cas de sortie de probabilités. [17]

4. Classification Des Images

L'apprentissage en profondeur dans les soins de santé donner un grand impact dans le monde , aussi par un prototype pour résoudre des problèmes médicaux. Il est incroyable de voir comment l'apprentissage en profondeur améliore le domaine de la santé, en particulier dans le diagnostic médical, et il améliorera la façon dont les médecins diagnostiquent et menacent les maladies.

Cette fois, la détection du parkinson dans les images de l'analyse des voix est une expérience formidable car le DL utilise des techniques avec la taille des données est importante, ces techniques doivent disposer d'une infrastructure haut de gamme pour s'entraîner dans un délai raisonnable. (J. Zheng et al., 2017). L'apprentissage en profondeur brille vraiment lorsqu'il s'agit de problèmes complexes tels que la classification d'images grâce à l'ensemble des caractéristiques trouver dans ces images. Dans notre travail nous montrons deux façons pour détecter la maladie dans les images des signales de voix, la

[17] (https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc)

première le réseau neuronal convolutionnel CNN et la deuxième le réseau neuronal récurrent RNN avec Keras et Tensorflow.

Parmi les anciennes expériences sur les voix convent a des images en deep Learning l'article de la détection du stéganographie MP3 dans le domaine du code entropique. Ces méthodes stéganographiques cachent des messages secrets dans le flux audio compressé grâce à la substitution de code Huffman, qui permet généralement d'obtenir une capacité élevée, une bonne sécurité et une faible complexité de calcul. Ces derniers basés sur CNN, la matrice de coefficients DCT modifiés quantifiés (QMDCT) est sélectionnée comme données d'entrée du réseau proposé. Pour les trois algorithmes de stéganographie, la précision de détection du réseau proposé est supérieure aux fonctionnalités artisanales, HCM-GEO de précision 78%, HCM-YAN de précision 85%, EECS de précision 90%, pour la taille arbitraire 128, HCM-GEO de précision 80 %, HCM-YAN de précision 90%, EECS de précision 95% pour la taille arbitraire 320.(Deng & Yu, 2013)

Dans cet article, un nouveau paradigme pour la stéganalyse basé sur le concept d'apprentissage des fonctionnalités. Il est applique la méthode (CNN) est utilisé pour apprendre automatiquement les fonctionnalités de la stéganalyse. Pour que CNN fonctionne mieux pour la stéganalyse, Ils proposent en outre d'utiliser une combinaison de modèles pour augmenter les performances de la méthode basée sur CNN. De plus, une stratégie de recadrage est proposée pour permettre un modèle de traiter des tailles d'image d'entrée arbitraires. Ils démontrent l'efficacité de la méthode proposée contre les algorithmes stéganographiques du domaine spatial de pointe tels qu'HUGO, WOW, SUNIWARD, MIPOD et HILL-CMD. Pour aider à comprendre les fonctionnalités apprises de CNN, ils fournit des visualisations des filtres appris et des cartes d'entités et une analyse quantitative des caractéristiques apprises des couches convolutives .(Qian et al., 2018)

5. Travaux récents utilisant Deep Learning pour la parole

Des réseaux de neurones récurrents et des réseaux avec des modèles acoustiques à convolution ont été explorés également utilisés dans la reconnaissance vocale il y a plus de 20 ans. Les réseaux convolutifs se sont également révélés bénéfiques pour le diagnostic des modèles acoustiques de la maladie de Parkinson à partir de la parole continue (Frid et al., 2017). Les réseaux de convolution profonds se sont concentrés sur l'automatisation du processus de diagnostic à partir de la parole native continue en supprimant la nécessité d'un personnel formé à partir du processus de diagnostic. Cela a été fait en utilisant une adaptation de l'architecture CNN pour le traitement du signal unidimensionnel (c'est-à-dire le signal vocal brut) sur un ensemble d'apprentissage relativement petit. Et pour un modèle profond d'architecture de réseau neuronal récurrent RNN qui apprend directement la similitude des patients à partir des dossiers des patients longitudinaux et multimodaux a été proposé dans (Che et al., 2017), qui apprend la similitude entre deux séquences d'enregistrements de patients longitudinaux grâce à une correspondance temporelle dynamique modèles dans les séquences des patients. Les évaluations des dossiers de patients réels démontrent l'utilité et l'efficacité prometteuses de l'architecture proposée dans les prévisions personnalisées.

Plus récemment, les DNN sont devenus un incontournable de la reconnaissance automatique de la parole ASR et ont été utilisés dans (Harar et al., 2018) pour la détection de pathologie vocale où ils ont utilisé des enregistrements vocaux de voyelle soutenue / a /

produits à hauteur normale à partir du corpus allemand Saarbruecken Voice Database (SVD). L'idée derrière cette expérience est l'utilisation de couches convolutives en combinaison avec des couches récurrentes de mémoire à long terme (LSTM) sur un signal audio brut. Chaque enregistrement a été divisé en segments fenêtrés de Hamming de 64 ms avec un chevauchement de 30 ms. Le modèle formé a atteint une précision de 71,36% avec une sensibilité de 65,04% et une spécificité de 77,67% sur 206 fichiers de validation et une précision de 68,08% avec une sensibilité de 66,75% et une spécificité de 77,89% sur 874 fichiers de test.(Al-Fatlawi et al., 2016). Il s'agit d'un résultat prometteur en faveur de l'approche, car elle est comparable à une expérience similaire publiée précédemment qui utilisait une méthodologie différente. Un réseau de croyances profondes (DBN) a également été adopté en tant que technique efficace pour diagnostiquer la maladie de Parkinson (PD) (Daphne, 2018). Le système proposé a été formé et testé avec les voix d'un certain nombre de patients et des personnes en bonne santé où les caractéristiques traditionnelles ont été extraites des signaux vocaux pour être entrées dans le réseau de croyances profondes (DBN) qui est utilisé pour créer un modèle de correspondance des voix. Dans cet article, le DBN a été utilisé pour classer la maladie de Parkinson qui comprend deux machines Boltzmann restreintes empilées (RBM) et une couche de sortie. Deux étapes d'apprentissage devaient être appliquées pour optimiser les paramètres des réseaux. La première étape a été un apprentissage non supervisé qui utilise des RBM pour surmonter le problème qui peut être dû à la valeur aléatoire des poids initiaux. Deuxièmement, l'algorithme de rétro propagation a été utilisé comme apprentissage supervisé pour le réglage fin. La précision globale des tests du système proposé était de 94%.(Shahi & Pant, 2018).

6. Conclusion

Dans ce chapitre on a présenté l'analyse acoustique en basant sur les paramètres MFCCs et comment on les a utilisés pour l'extraction des images. Ensuite, l'architectures des réseaux de neurones les plus connues et les plus utilisées ainsi que les nouvelles architectures qui semblent avoir un avenir prometteur dans différents domaines d'application de la technologie contemporaine, avec certain méthodes d'évaluation ont été présenté. Ensuite, nous avons parlé sur les travaux récents liés à- la pathologie qui les ont utilisés et de certaines applications propulsées par eux dans différents domaines.

Chapitre III

Approche De Travail

Table des matières

1.Introduction 73
2.Approche proposée 73
2.1. Schéma du Modèle proposé 73
3.Base de données 75
3.1. Définition (généralité) 75
4.Extraction des paramètres MFCCs 75
5.Classification 76
5.1 Définition (généralité) 76
5.2 Méthode proposée pour la classification des images 76
5.3 Architecture du Modèle CNN proposé 77
5.4 Architecture du Modèle Apprentissage par transfert proposé 79
6.Conclusion 82

Liste des Figures

Figure III.1 : Schéma du Modèle proposé. 68
Figure III.2 : modèle VGG16 73
Figure III.3 : Modèle InceptionV3 75
Figure III.4 : Modèle ResNet 76

Liste des Tableaux

Tableau III.1 : L'architecteur de notre réseau CNN..... 72

1. Introduction

Dans ce chapitre, nous présentons notre approche de travail et les modèles utilisés pour la classification des images. Une méthode de détection de la maladie de Parkinson en s'appuyant sur la conversion d'une base de données audio en un ensemble de données contenant des images de cartes sonores va être expérimentée. Ensuite, ces images vont être utilisées pour une détection d'une maladie de parkinson comme étude de cas par une classification basée sur l'apprentissage profond.

2. Approche proposée

La première étape est de présenter les différentes phases suivies pour concevoir notre approche proposée, nous expliquons les détails où nous prétraitons les données, puis utilisons différentes méthodes sur les données pour améliorer les performances de notre modèle de DL (CNN).

2.1 Schéma du Modèle proposé

En raison du manque de capacités et de données nécessaires dans les domaines de l'analyse et de la classification des sons. Dans ce simple travail, nous avons proposé méthode d'analyse et de classification des sons, en particulier dans le domaine de la médecine. Cette méthode cherche à réduire les difficultés en convertissant le son en image et en travaillant dessus. A obtenu des résultats étonnants dans le domaine des images. Pour la concrétisation de notre approche, nous avons adopté le schéma présenté dans la Figure III.4 du modèle de la méthode de conception. La première étape consiste à définir une base de données dont on va sélectionner l'ensemble d'enregistrements utilisés. L'extraction des paramètres acoustiques vient en deuxième étape. La transformation des ces paramètre en images est présentée dans la quatrième phase. Les performances de la méthode utilisé est mesurée par un model en utilisant trois types de classificateur de apprentissages par transfert comme un type d'un réseau de neurones convolutifs (CNN).

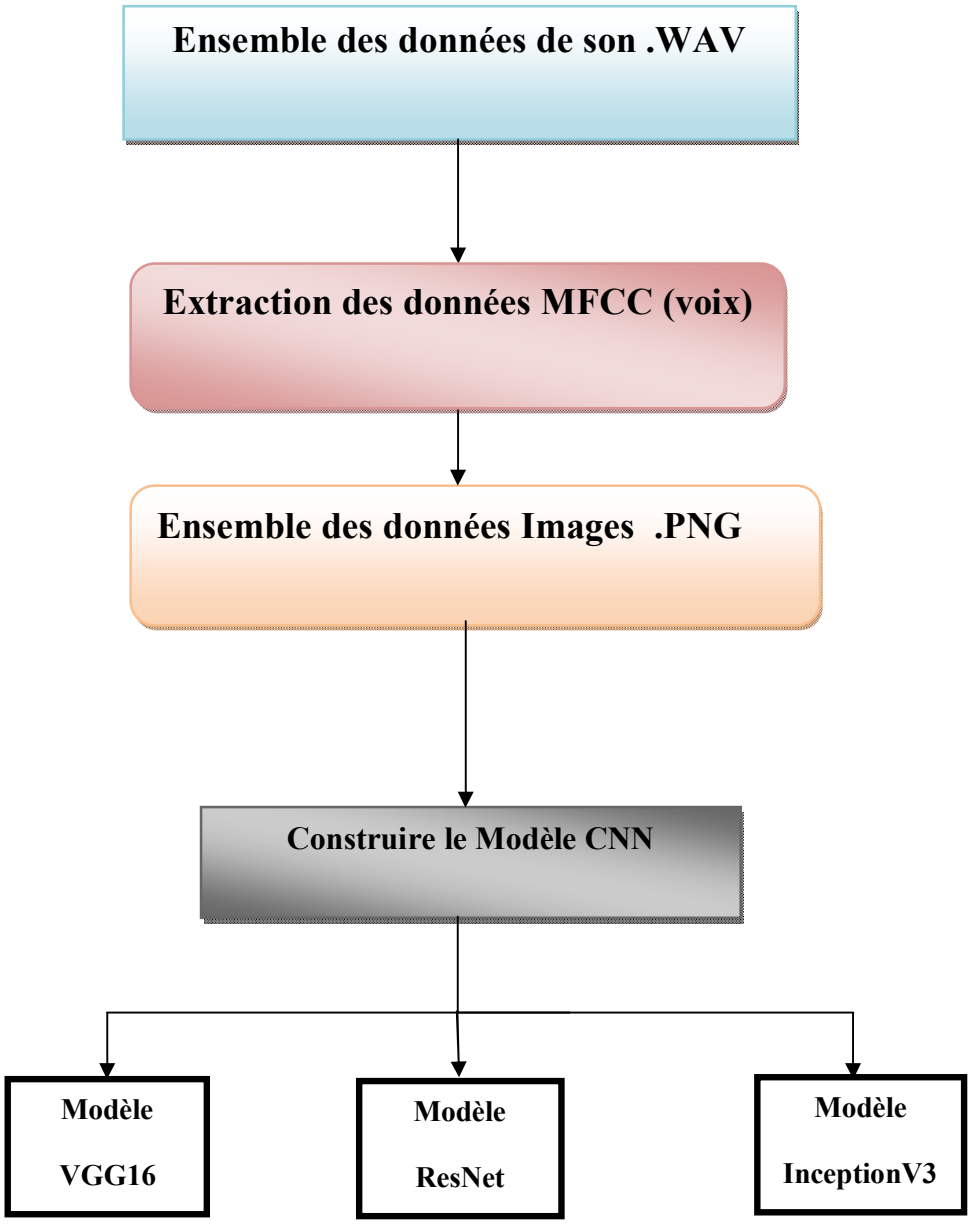


Figure III.1 : Schéma du Modèle proposé.

3. Base de données

3.1 Définition (généralité)

Une base de données est une collection d'informations organisées afin d'être facilement consultables, gérables et mises à jour[18]. Au sein d'une database, les données sont organisées dans un format prêt à être traité. Elles sont indexées afin de pouvoir facilement trouver les informations recherchées à l'aide d'un logiciel informatique. Chaque fois que de nouvelles informations sont ajoutées, les données sont mises à jour, et éventuellement supprimées (Akujuobi & Zhang, 2001.). Les bases de données sont stockées sous forme de fichiers ou d'ensemble de fichiers sur un disque magnétique, une cassette, un disque optique ou tout autre type d'appareil de stockage(Beel et al., 2017). Les bases de données sont principalement présentes au sein des systèmes mainframe les plus larges, mais on les retrouve aussi au sein de workstations distribuées de moindre envergure et autres systèmes midrange comme les IBM AS/400 ou même les ordinateurs personnels. (Hachesu et al., 2017) Les données de la base peut être organisé d'une manière structuré (Tableau, linges, colonnes) ou plus complexe lorsque les données sont dans un format non structuré tel que l'image ou l'audio, elles ne peuvent pas être utilisées directement pour le traitement et elles doivent être présentées de manière standard pour utiliser.

4. Extraction des Images MFCCs

L'extraction de la meilleure représentation paramétrique des signaux acoustiques est une tâche essentielle dans la conception de tout système de reconnaissance vocale. Les Mel-Frequency Cepstral Coefficients (MFCCs) est l'une des techniques d'extraction de caractéristiques les plus populaires utilisées dans la reconnaissance vocale, ce qui signifie qu'il extrait des paramètres de la parole similaires à ceux qui sont utilisés par les humains pour entendre la parole.

Dans notre approche nous utilisons ces les coefficients MFCCs pour extraire les images de notre base de données. Les MFCCs basé sur la transformation linéaire en cosinus d'un spectre de puissance logarithmique sur une échelle de fréquence logarithmique Mel non linéaire. L'extraire des images MFCCs se fait par l'application de préaccentuation, puis l'analyse par transformée de Fourier est effectuée à l'aide de fenêtres de Hamming. Ensuite, le spectre de puissance basé sur le périodogramme est estimé pour chaque trame segmentée. L'ensemble de bancs de filtres à l'échelle de Mel est calculé et appliqué au spectre de puissance de chaque cadre. Pour calculer le spectre d'énergie du banc de filtre, chaque filtre est multiplié par le spectre de puissance et les coefficients sont additionnés. La sortie de bancs de filtres est transformée en une échelle logarithmique. Finalement, la transformée en cosinus discrète (DCT) est appliquée sur les amplitudes logarithmique du banc de filtre pour obtenir les vecteurs caractéristiques MFCC. Ces vecteurs sera plote dans des images différant, chaque une selon les valeurs de ces vecteurs.

[18] https://fr.wikipedia.org/wiki/DATASET

La voix d'une personne malade contient divers troubles qui sont clairement prononcés par la parole, par rapport à la personne non malade. Ce changement de ton de la voix se manifeste sous la forme de certaines différences et changements qui diffèrent d'une personne à l'autre par rapport au ton de la personne Non infecté. La nette différence dans les caractéristiques du son entre la personne atteinte de la maladie de Parkinson et les personnes non infectées, nous a permis d'expérimenter la méthode de détection de la maladie à travers des images, en s'appuyant sur la conversion du ton de la voix en une image. Après, la préparation de la base de données qui est une ensemble des images contient un graphe de son la voix, ces images sont organisées en deux répertoires « training data », pour l'apprentissage de modèle et « «test data » pour le test de classifieur. Chaque répertoire contient deux sous-répertoires « HI» et « PWP » pour différentes classes (malade et non malade). À l'intérieur de chaque sous-répertoire, les images sont organisées par un numéro de son de chaque section.

5. Classification

5.1 Définition (généralité)

La classification est une méthode qui sert a considéré une hiérarchie des classes, et classer les échantillons dans les classes du niveau ancêtre. Le principe de la classification est :

- Descendre dans la hiérarchie pour classer les échantillons dans les classes du niveau courant. Le niveau suivant sera celui atteint en passant par la classe prédite dans le niveau actuel.
- Itérer jusqu'à arriver au niveau le plus bas de la hiérarchie.
- Il existe toutefois, des méthodes qui effectue la classification dans le sens inverse, en commençant par les classes du niveau le plus bas puis continuent le traitement en allant vers les niveaux de haut: Down-Top stratégie.

5.2 Méthode proposée pour la classification des images

L'un des problèmes les plus difficiles du traitement de la parole est la classification de la voix. Bien que de nombreuses études axées sur l'extraction de caractéristiques et la conception de classificateurs à améliorer aient été menées, la précision de la classification est quelque peu difficile en raison d'un manque d'informations et de données suffisantes pour cela. Le principal problème avec l'identification de la voix d'un patient est que la voix est influencée par d'autres facteurs qui rendent difficile la création de fonctionnalités puissantes et la conception d'un classificateur en profondeur. Alors que, dans de nombreux cas, les informations audio dans le discours de l'orateur disparaissent en étant affectées par de nombreux facteurs tels que le bruit de fond, le contenu de la parole et les différences audio.

Le succès de l'architecture d'apprentissage en profondeur DL dans de nombreuses applications de ce travail a stimulé la proposition d'un système qui pourrait classer le son normal d'anormal. Les signaux phonographiques contiennent des informations très utiles sur l'état de la voix, en particulier pour les maladies affectant les cordes vocales. Et cette méthode d'enregistrement des sons peut être représentée sur le graphique en appliquant certaines caractéristiques, notamment MFCC. Et en analysant ces signaux, la détection précoce de la maladie de Parkinson (qui affecte également une grande partie de la voix) peut être détectée. Par conséquent, l'analyse intelligente et automatisée du diagramme vocal est extrêmement

importante pour déterminer si la voix d'une personne fonctionne correctement ou si elle doit être référée à un expert pour une évaluation plus approfondie.

L'idée principale est de traduire le signal audio en une image 2D représentant les coefficients cepstraux temporels et la fréquence Mel. Pour ce faire, nous développons une nouvelle approche qui utilise cette fonctionnalité. Où nous divisons d'abord les signaux de carte audio qui sont des images et utilisons un algorithme basé sur le modèle CNN et extrayons un ensemble de caractéristiques du domaine temporel et fréquentiel du tracé sonore. Le modèle distingue alors les signaux phonographiques normaux et malades.

5.3 Architecture du Modèle CNN proposé

Notre travail a pour objectif de classifier des images composés des graphiques de la voix d'une personne et nous ne devons distinguer que deux classes: si ce personne malade ou non dans une image.

Un réseau CNN est construit. Les images transformées des coefficients MFCCs sont mis comme entrée de model CNN. Ces images sont bien couleur composer d'un graphe en bleu avec l'arrière plan en blanc (pas noire et blanc), de taille 120x120 dans notre convnet. Ensuite, nous empilerons 3 modules {convolution + relu + maxpool}. Notre CNN avec trois couches de convolution / pooling et deux couches entièrement connectées. Les trois couches conv utiliseront respectivement 32, 32 et 64 filtres 3x3. J'ai utilisé des abandons sur les deux couches entièrement connectées pour éviter le sur-ajustement. Nos convolutions opèrent sur des fenêtres 3x3 et nos couches maxpool fonctionnent sur des fenêtres 2x2. Nous collons deux couches entièrement connectées. Parce que nous sommes confrontés à un problème de classification à deux classes (classification binaire), nous terminerons notre réseau par une activation relu, de sorte que la sortie de notre réseau sera un seul scalaire entre 0 et 1, codant la probabilité que le courant l'image est de classe 1 (par opposition à la classe 0). Les couches de convolution réduisent légèrement la taille des cartes d'entités en raison du remplissage, et chaque couche de mise en commun divise par deux la carte d'entités. Ensuite, nous allons configurer les spécifications pour la formation des modèles. Nous allons entraîner notre modèle avec la perte **binary_crossentropy loss**, car c'est un problème de classification binaire et notre activation finale est un sigmoïde. Nous utiliserons l'optimiseur de descente de gradient stochastique avec un taux d'apprentissage de 0,01 et un élan de 0,9. Pendant l'apprentissage, nous voudrons surveiller la précision de la classification. Et nous configurons des générateurs de données qui liront les images dans nos dossiers source, les convertiront en tenseurs float32 et les alimenteront (avec leurs étiquettes) sur notre réseau. Nous allons utiliser la classe ImageDataGenerator. Il permet de diffuser indéfiniment les images par lots à partir des dossiers train et validation. À l'intérieur de l'objet ImageDataGenerator, nous allons introduire des modifications aléatoires sur chaque lot. C'est un processus que nous appelons l'augmentation des données. Cela permet de générer plus de données afin que notre modèle ne voie jamais exactement deux fois la même image. Cela permet d'éviter le sur-ajustement et de mieux généraliser le modèle. Nous prétraiterons nos images en normalisant les valeurs des pixels pour qu'elles soient dans la plage [0, 1]. Rappelons qu'une image est un ensemble de pixels. Chaque pixel étant un triplet de valeurs (R, G, B), notre modèle prend en entrée une image de taille (120*120*3).

Model : « model CNN »		
Layer (type)	**Output Shape**	**Param #**
Input Layer	[(None, 120, 120, 3)]	0
Conv2d_1 (conv2D)	[(None, 120, 120, 32)]	896
Max_pooling2d_1 (maxPooling2)	[(None, 75, 75, 32)]	0
Conv2d_2 (conv2d)	[(None, 75, 75, 32)]	9248
Max_pooling2d_2 (maxPooling)	[(None, 37, 37, 32)]	0
Conv2d_3 (conv2D)	[(None, 37, 37, 64)]	18496
Max_pooling2d_3 (maxPooling)	[(None, 18, 18, 64)]	0
Dropout_1 (Dropout)	[(None, 18, 18, 64)]	0
Flatten_1 (Flatten)	[(None, 20736)]	0
Dense_1 (Dense)	[(None, 64)]	1327168
Dropout_2 (Dropout)	[(None, 64)]	0
Dense_2 (Dense)	[(None, 1)]	130
Totale params : 1, 355,938 Trainable params : 1, 355, 938 Non_trainable params : 0		

Tableau III.1 : L'architecteur de notre réseau CNN.

5.4 Architecture du Modèle Apprentissage par transfert proposé

Après le développement et les résultats avances dans ce domaine de classification des images par ces modèles pré-entraîne nous décidons de créer un nouveau modèle mis à jour qui montre comment utiliser ces réseaux de pointe dans notre projets de classification. Plus précisément, nous allons créer un script Python spécial pour classer nos images d'entrée personnalisées. Selon les modèles présentes au dessus

Premièrement nous appliquons le modèle VGG16 qu'est caractérisé par sa simplicité, il entre une image de taille 224*224*3. Ensuit il commence l'entrainement par 3 × 3 couches convolutives empilées les unes sur les autres de plus en plus en profondeur. La réduction de la taille du volume est gérée par la mise en commun maximale. Deux couches entièrement connectées, chacune avec 4 096 nœuds, sont ensuite suivies d'un classificateur softmax. Le 16 représentent le nombre de couches de poids dans le réseau. La figure III.2 présenté l'architecteur de notre modèle VGG16 implémenté.

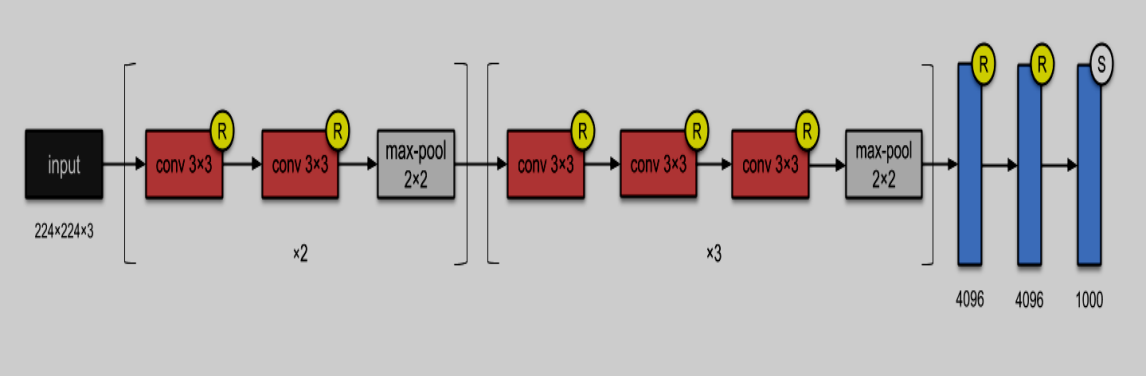
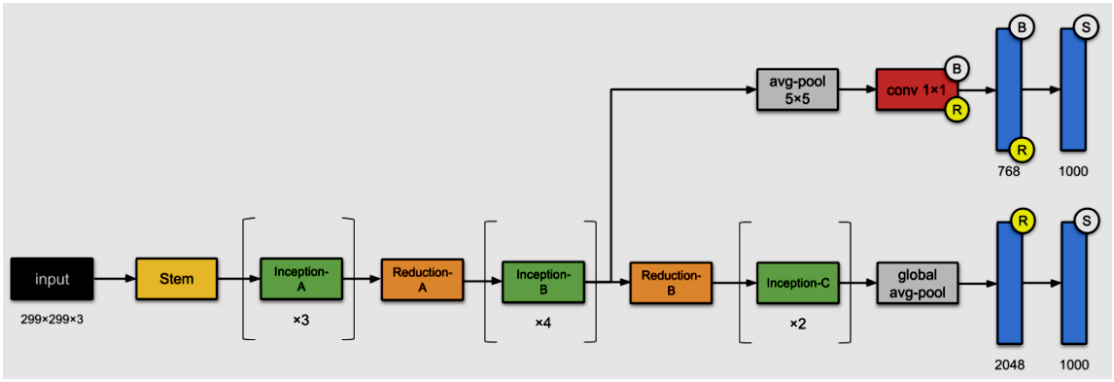
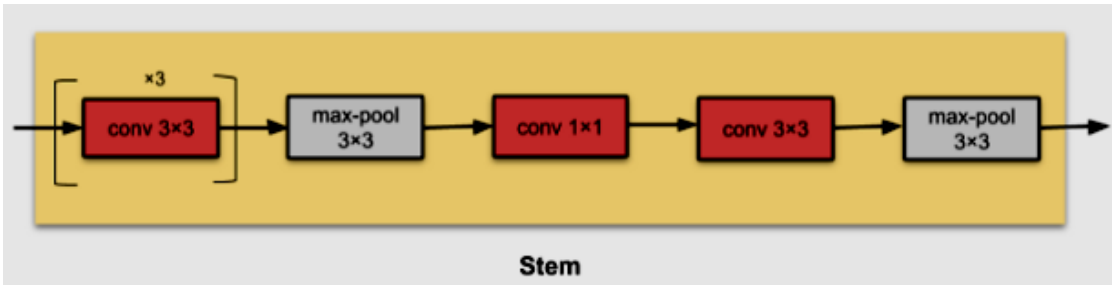


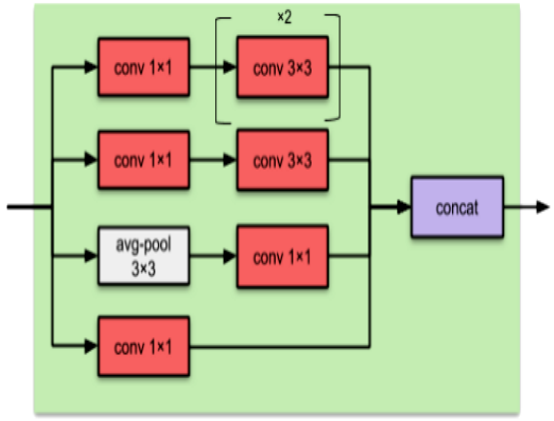
Figure III.2 : modèle VGG16

Ensuite, nous appliquons le modèle inceptionV3 prendre en entre une image de taille 299*299*3 cette image sera converti en tableau numpy de 0 et 1 afin de travailler avec keras, sera placées sur 434 grilles telles que produites par l'étape de prétraitement. Ce réseau utilisé en l'entrainement le calcul des convolutions 1 × 1, 3 × 3 et 5 × 5 dans le même module empilées les unes sur les autres de plus en plus en profondeur, la sortie de ces filtres est empilée le long la dimension du canal et avant d'être introduit dans la couche suivante du réseau, sont des couches de mise en commun et de filtres convolutifs avec des fonctions d'activation. Trois couches entièrement connectées de taille 1024, 512 et 3 sont ajoutées à la couche de concaténation finale. Un décrochage avec un taux de 0,6 est appliqué avant les couches entièrement connectées comme moyen de régularisation. Toutes les étapes présentées dans la figure III.3

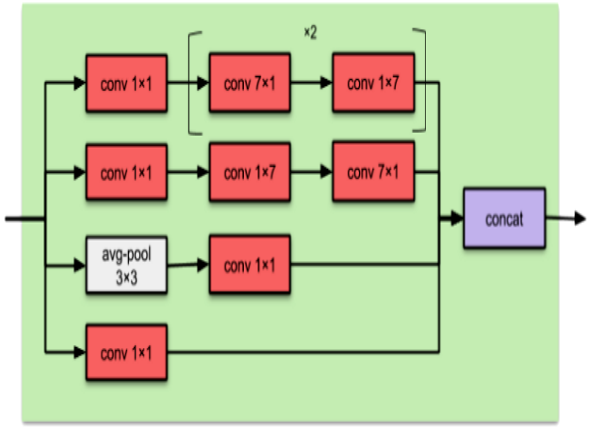




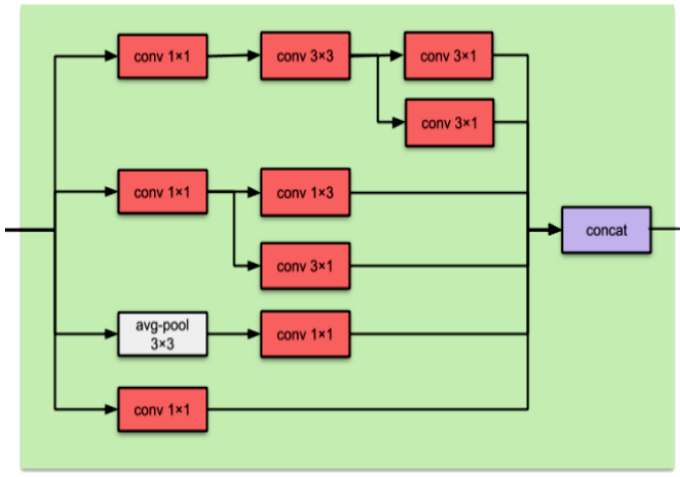
Stem



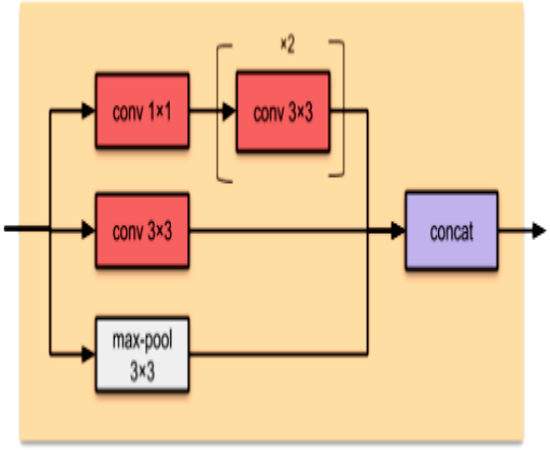
Inception-A



Inception-B



Inception-C



Reduction-A

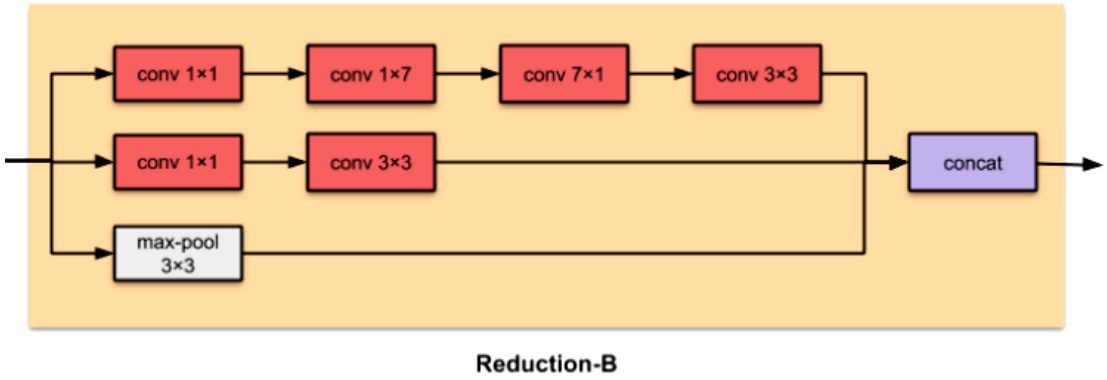
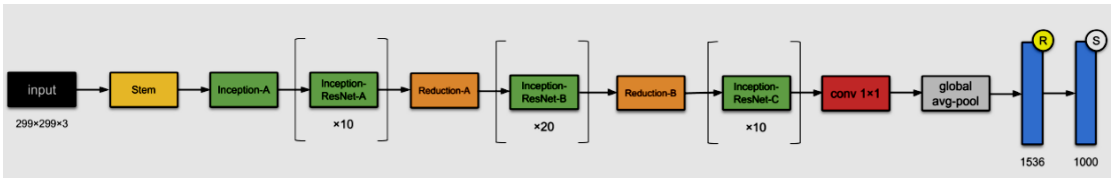
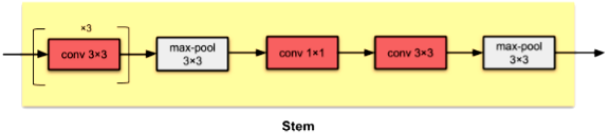
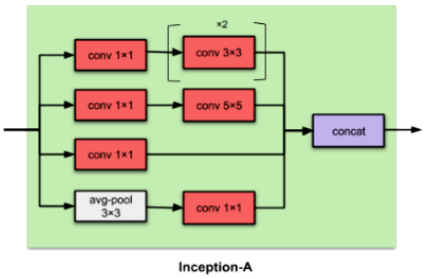


Figure III.3 : Modèle InceptionV3

Ensuit, nous avons appliquée le modèle ResNet sur notre ensemble des images qui entre une image de dimension 299*299*3 puis il commencer de faire l'entraînement sur les images telle que ResNet composer de 3×3 conception de couches convolutives. Le bloc résiduel a deux 3×3 couches convolutives avec le même nombre de canaux de sortie. Chaque couche convolutionnelle est suivie d'une couche de normalisation par lots et d'une fonction d'activation ReLU. Ensuite, nous sautons ces deux opérations de convolution et ajoutons l'entrée directement avant la fonction d'activation finale de ReLU. Ce type de conception nécessite que la sortie des deux couches convolutives ait la même forme que l'entrée, pour changer le nombre de canaux, nous devons introduire un 1×1 couche convolutive pour transformer l'entrée en la forme souhaitée pour l'opération d'addition. Ce modèle utilise trois modules constitués de blocs Inception dont chacun utilise plusieurs blocs résiduels avec le même nombre de voies de sortie. Le nombre de canaux dans le premier module est le même que le nombre de canaux d'entrée. Enfin, nous ajoutons les couches de pooling, suivie de la sortie de couche entièrement connectée, il y a 50 couches au total. Par conséquent, ce modèle est communément appelé ResNet-50. La figure III.4 présenté l'architecteur de ce modèle.







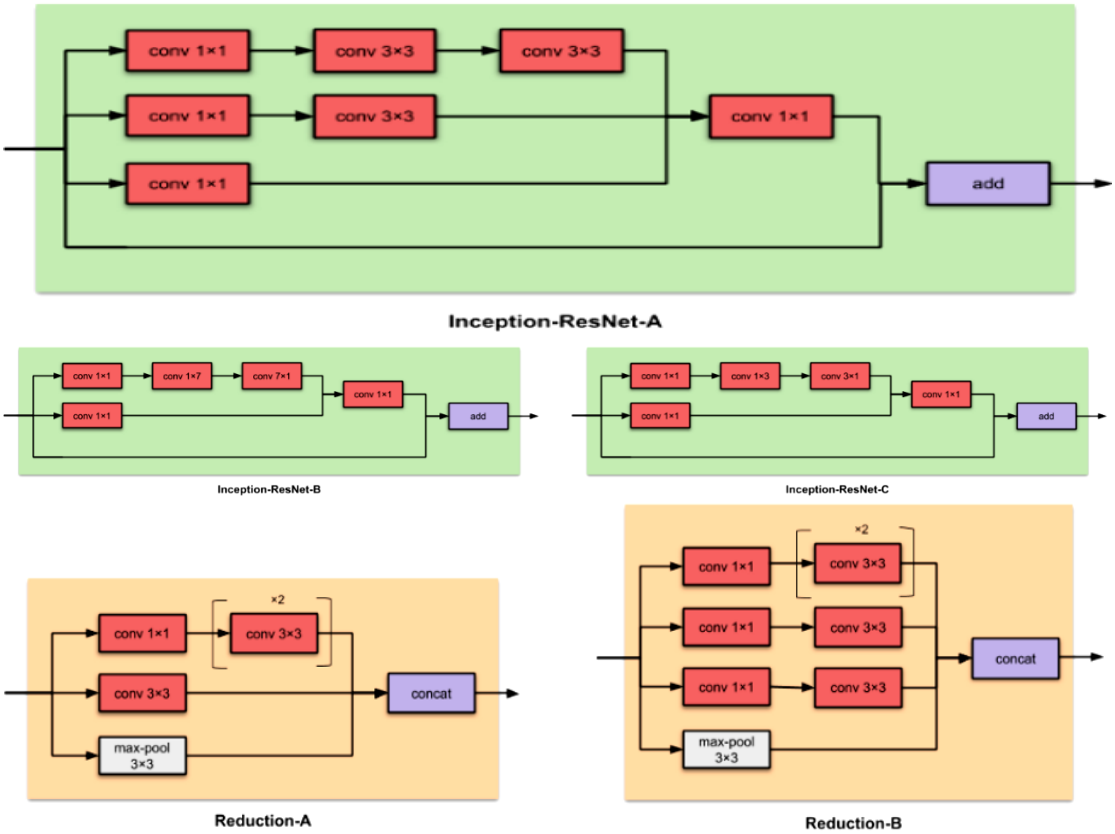


Figure III.4 : Modèle ResNet

6. Conclusion

Dans ce chapitre, nous avons présenté une approche de classification des images qui cherche à concevoir un modèle pour la détection de la pathologie vocale «cas de maladie de Parkinson» basée sur les réseaux de neurones convolutionnels. Et nous avons présenté en détail toutes les étapes de la méthode proposée, et les principaux algorithmes composant dans notre travail.

Chapitre IV

Test Et Résultats

Table des matières

1.Introduction 87
2.Les outils et les librairies utilisés 87
2.1. TensorFlow.....87
2.2 Keras.....87
2.3. Python.....88
2.4 Configuration utilisé dans l'implémentation..... 88
3. Description de la Base de données 89
3.1 Base de données originale..... 89
3.2 La modélisation des fichiers sons en images.....90
4.Expérimentations et résultats 93
4.1 Résultat obtenue pour modèle CNN 93
4.2 Paramètre d'évaluation du réseau CNN.....94
4.3 Résultat obtenue de Modèle apprentissage par transfert.....94
4.4 L'évaluation 96
5.Conclusion 98

Liste des Figures

Figure IV.1 : Logo de TensorFlow 81
Figure IV.2 : Logo de Keras 82
Figure IV.3 : Logo de Python 82
Figure IV.4 : Graphique de mot "ketap" d'un patient non malade 84
Figure IV.5 : Graphiques 20_1814, 20_1829, 20_1834 homme non malade 85
Figure IV.6 : Graphique de mot "ketap" d'un patient malade 85
Figure IV.7 : graphique 0020_14174, 0020_15101, 0020_16123 pour femme malade 85
Figure IV.8 : Apprentissage et Validation modèle CNN 88
Figure IV.9 : matrice de confusion de notre CNN 89
Figure IV.10 : Apprentissage et Validation modèle VGG16 89
Figure IV.11 : Apprentissage et Validation modèle ResNet 91
Figure IV.12 : Apprentissage et Validation modèle InceptionV3 91
Figure IV.13 : Matrice de confusion InceptinV3 92
Figure IV.14 : Matrice de confusion ResNet 92
Figure IV.15: Matrice de confusion VGG16 93

Liste des Tableau

Tableau IV.1 : les bibliothèques python utilisé 83
Tableau IV.2 : Accuracy des Modèles 91
Tableau IV.3 : Rapport de classification des modèles 91

1. Introduction

Dans ce chapitre, nous présentons les différents outils et les librairies utilisés du modèle proposé ainsi que son implémentation. Pour entrainer le modèle nous allons exposer les caractéristiques de la base de données puis nous allons discuter les résultats obtenus.

2. Les outils et les librairies utilisés

2.1. TensorFlow

TensorFlow est une bibliothèque open source de Machine Learning, Créé par l'équipe Google Brain en 2011, sous la forme d'un système propriétaire dédié aux réseaux de neurones de Deep Learning, pour le calcul numérique utilisant des graphes de flux de données. Les nœuds des graphes représentent des opérations mathématiques, tandis que les arêtes des graphes représentent les tableaux de données multidimensionnels (tenseurs) qui circulent entre eux. Cette architecture flexible vous permet de déployer des calculs sur un ou plusieurs processeurs ou un processeur graphique (Graphics Processing Unit) sur un ordinateur de bureau, un serveur ou un périphérique mobile sans réécrire le code. Cette bibliothèque permet notamment d'entraîner et d'exécuter des réseaux de neurones pour la classification des chiffres écrits à la main, la reconnaissance d'image, les plongements de mots, les réseaux de neurones récurrents, les modèles sequence-to-sequence pour la traduction automatique, ou encore le traitement naturel du langage.(Deep & Tensorflow, 2018)



Figure IV.1 : Logo de TensorFlow

2.2. Keras

Keras c'est une API de haut niveau permettant de créer et de former des modèles d'apprentissage en profondeur, capable de s'exécuter sur Tensorflow, il est utilisé pour le prototypage rapide, la recherche avancée et la production. Les modèles Keras sont fabriqués en reliant des blocs de construction configurables ensemble, donc Keras nous permet de créer de nouvelles couches, des fonctions et développer des modèles à la pointe de la technologie avec peu de restrictions en quelques lignes de code.[19]

[19] https://keras.io/



Figure IV.2 : Logo de Keras

2.3. Python

Python est un langage objet interprété de haut niveau, créé au début des années quatre-vingt-dix par Guido Van Rossum au Centrum voor Wiskunde à Informatisa, Amsterdam Python est un langage de programmation facile à utiliser et puissant. Il offre des structures de données puissantes de haut niveau et une approche simple mais réelle de la programmation orientée-objet. La syntaxe élégante de python et le typage dynamique, ajoutés à sa nature interprétée, en font un langage idéal pour écrire des scripts et pour le développement rapide d'applications dans de nombreux domaines et sur la plupart des plates-formes.(Hopper, 1977)



Figure IV.3 : Logo de Python

2.4. Configuration utilisé dans l'implémentation

La configuration du matériel utilisé dans notre implémentation est :

- Un pc bureau Lenovo i5 CPU 510 GHZ
- RAM de taille 8 GO
- Système d'exploitation Windows 7 64 bit

La configuration logicielle utilisée pour notre développement est fait sur Anaconda qui est doté de Python 3.7 ainsi que l'éditeur jupter 3.3.4 qui a été choisi pour sa simplicité et efficacité. Plusieurs bibliothèques ont été utilisées : Tensorflow/keras, scipy, numpy, pandas, matplotlib, sklearn.

Bibliothèque	Version utilisée
Keras	2.2.4
Tensorflow	1.14.0
Matplolib	3.0.3
Numpy	1.16.3
Pandas	0.24.2
Scipy	1.2.1
Sklearn	0.21.1
System (python)	3.7.3 (default, mar 20 2019, 08:33:11) [MSC v.1915 64 bit (AMD64)]

Tableau IV.1 : les bibliothèques python utilisé

3. Description de la Base de données

3.1 Base de données originale

Généralement, un ensemble de base de données est collecté dans un but spécifique et à partir de diverses sources. Mais parfois, l'ensemble des informations de la base peut être plus complexe lorsque les données sont dans un format non structuré tel que l'image ou l'audio et elles ne peuvent pas être utilisées directement pour le traitement et elles doivent être présentées de manière standard pour être plus approprié pour l'analyse. Dans notre travail, nous avons utilisé un ensemble de données sur la parole de Parkinson avec plusieurs types d'enregistrements sonores.

Les voix des personnes représentant la base de données sont celles qui ont fait appel au Département de neurologie de la faculté de médecine de Cerrahpasa, à l'Université d'Istanbul (Isenkul, Sakar et Kursun, 2014) et (Sakar et al., 2013). Plusieurs types d'enregistrements audio sont réalisés sur des femmes, et des hommes et de tous les sujets (environ 26 échantillons de phonèmes comprenant des voyelles continues, des nombres, des mots et des phrases courtes.

L'ensemble de données des sons comprend 1040 échantillons sont divisés en deux type de classes. 520 individu sont dans la classe malade qu'elle nommée « PWP » et les 520 restons sont classés dans la classe non malade « HI ».

3.2 La modélisation des fichiers sons en images

La méthode utilisée pour garantir ce transfert des données c'est l'extraction de caractéristiques MFCCs. Cette phase est fondamentale du système de reconnaissance audio, préalable à la classification. Dans notre approche nous appliquons cette méthode sur notre

ensemble de données audio (1040 sons), et nous extraire l'ensemble des images (1040 images correspondantes aux nombres des fichiers sons). Notre objectif dans cette approche et de augmenter le nombre de données pour utiliser dans le deep learning, donc nous avons consisté à disposer d'une grande base de données, suffisamment large pour garantir une formation spécifique au système afin de le rendre le plus efficace possible dans le diagnostiquer de la maladie en question. Notre nouvelle base de données c'est un ensemble des images, ces derniers composé des nombre de courbes pour présenter la parole d'un patient. Dans notre approche nous allons augmenter la base grâce à la séparation de les courbes, chaque courbe devient une image qui représenter un état malade ou non. Dans ce programme nous appliquons premièrement, la lecture des fichiers audio et fais le calcule des coefficients de MEL, (cette fonction a besoin certaines entrées pour établir ces calcules). En résultat, les coefficients sont présentés dans une matrice (sachant que le nombre des lignes est fixé comme argument), mais la deuxième dimension se varie à dépend au contenu du fichier sons. Cette matrice est ajoutée à la structure de données qui contient toute les matrices des fichiers audio. Les graphes qui font les images pour le dataset sont tracés ensemble pour chaque matrice. Mais pour étendoir le dataset, on va tracer chaque courbe dans une image seul, telle que au lieu d'avoir une image contenant beaucoup de coubre pour chaque fichier audio, on aura donc des centaines de courbes pour chacun. Comme l'exemple suivante, le mot "ketap" prend le numéro "18" dans les enregistrements audio dans la base des audois, le fichier 20_18 pour l'homme, du matrice des coefficients de 13x39, c-à-d : 39 courbes, comme le montre la figure IV.4

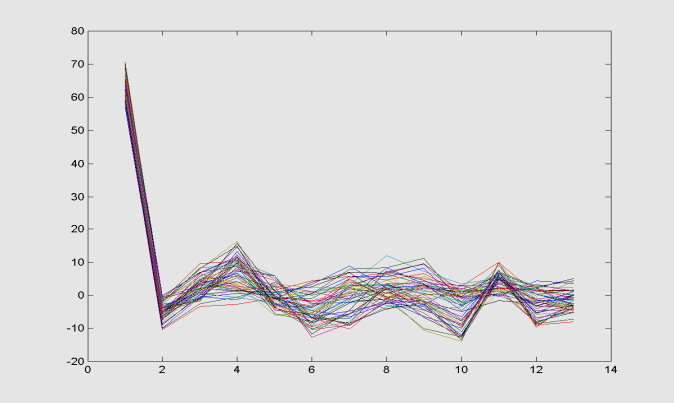


Figure IV.4 : Graphique de mot "ketap" d'un patient non malade

Après nous allons faire la séparation de ces courbes et on obtient N images dont chaque une représenter un courbe. La valeur de N sera variée tous dépend d'un échantillon a un autre et selon sa durée. La figure IV.5 représenter une parmi les images représentatives.

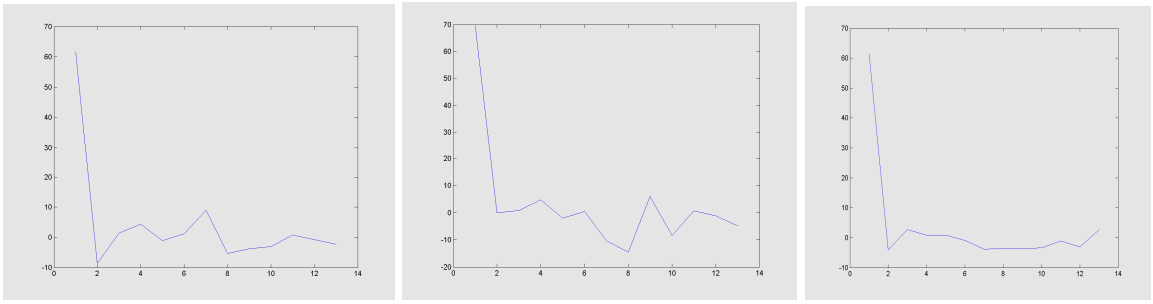


Figure IV.5 : Graphiques 20_1814, 20_1829, 20_1834 homme non malade

Aussi le fichier 0020_18 pour la femme, de la matrice des coefficients de 13x35, c-à-d : 35 courbes, comme le montre dans la figure IV.6

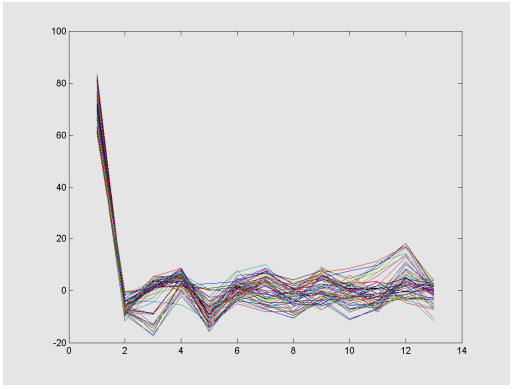


Figure IV.6 : Graphique de mot "ketap" d'un patient malade

Cette image donne 35 courbes dont les trois premières courbes sont présentées dans la figure IV.7

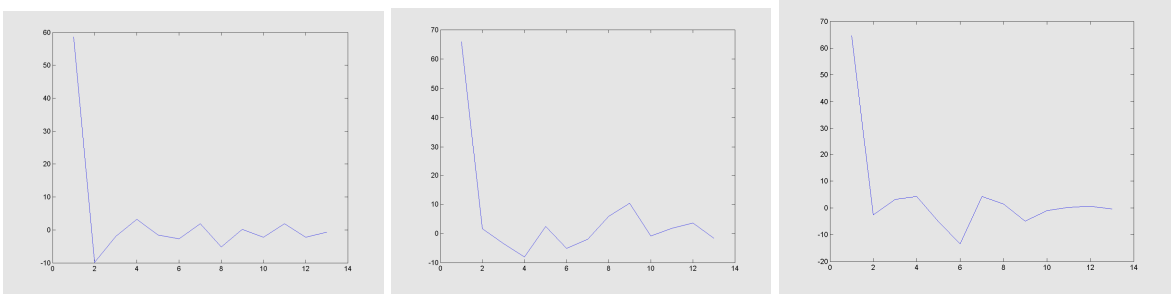


Figure IV.7 : graphique 0020_14174, 0020_15101, 0020_16123 pour femme malade

Et le fichier 0019_18 pour l'homme, du matrice des coefficients de 13x52, c-à-d : 52 courbes, et le fichier 19_18 pour la femme, du matrice des coefficients de 13x34, c-à-d : 34 courbes....et ainsi de suite, avec 18_18, et 0018_18, 17_18 et 0017_18 ...etc.

Après le parcours de tous les fichiers audio, nous allons crée 43 363 images pour la classe HI, et 47 219 pour la classe PWP, totalement 90 582 images bien classées et distribuées selon les

critères nécessaires à faciliter notre étude. Les principes étapes pour l'extraction des images sont :

MFCC (S, FS, TW, TS, ALPHA, WINDOW, R, M, N, L) renvoie des coefficients cepstraux de fréquence mel (MFCC) calculés à partir du signal de parole donné dans le vecteur S et échantillonné à FS (Hz).

Le signal de parole est d'abord pré accentué en utilisant un filtre FIR de premier ordre avec un coefficient de préaccentuation ALPHA.

Le signal de parole pré accentué est soumis à l'analyse par transformée de Fourier de courte durée avec des durées de trame de TW (ms), des décalages de trame de TS (ms) et une fonction de fenêtre d'analyse donnée comme poignée de fonction dans WINDOW.

Ceci est suivi d'un calcul du spectre de magnitude suivi d'une conception de banc de filtres avec M filtres triangulaires uniformément espacés sur l'échelle de mel entre les limites de fréquence inférieure et supérieure données en R (Hz).

Entrées :

- **S** est le signal vocal d'entrée (en tant que vecteur)
- **FS** est la fréquence d'échantillonnage (Hz)
- **TW** est la durée de la trame d'analyse (ms) (Tw = 25;)
- **TS** est le décalage de trame d'analyse (ms) (Ts = 10;)
- **ALPHA** est le coefficient de préaccentuation (alpha = 0.97;)
- **WINDOW** est un handle de fonction de fenêtre d'analyse (window = 'hamming';)
- **R** est la plage de fréquences (Hz) pour l'analyse du banc de filtres (R = [300 3700];)
- **M** est le nombre de canaux du banc de filtres (M = 20;)
- **N** est le nombre de coefficients cepstraux (C = 13;)
- **L** est le paramètre de levage (L = 22;)

Les sorties

- **CC** est une matrice de coefficients cepstraux de fréquence mel (MFCC) avec des vecteurs de caractéristiques sous forme de colonnes
- **FBE** est une matrice des énergies du banc de filtres avec des vecteurs de caractéristiques sous forme de colonnes.
- **FRAMES** est une matrice de cadres fenêtrés (une image par colonne).

4. Expérimentations et résultats

4.1 Résultat obtenue pour modèle CNN

Pour reconnaitre les images des sons malades et non malades nous basons sur les modèles précédents (présenter en chapitre III), en utilisant des réseaux de neurones convolutionel. Le premier élément à prendre en compte est le choix d'architecture du CNN. Les architectures de réseaux neuronaux sont des combinaisons des couches qui se

transmettent les sorties les unes après les autres. Nous avons choisi une architecture initiale en fonction des approches rencontrées lors de notre étude. Des résultats de performance sont obtenus en fonction du choix et de la séquence des couches constituant les architectures de réseau neuronal. Nous configurons les variables à utiliser lors du prétraitement de l'ensemble de données et de la formation du réseau. Puis nous formons les images en tenseurs à virgule flottante prétraités de manière appropriée avant de les alimenter sur le réseau. Nous décodons le contenu de ces images au format de grille approprié selon leur contenu RGB, et nous redimensionnons les tenseurs de valeurs comprises entre 0 et 255 à des valeurs comprises entre 0 et 1, grâce a la classe ImageDataGenerator. Ensuite nous appliquons le model et nous choisissons l'optimiseur ADAM et la fonction de perte d'entropie croisée binaire. Nous avons configuré un taux d'apprentissage de 0,001. Nos données ont été divisées en 80% pour l'apprentissage et 20% pour le test. Nous avons pris 20% du train comme validation, et il est courant d'entraîner le modèle pendant plusieurs périodes et un ensemble d'échantillons d'entrée. Avec un grand nombre d'époques pour éviter le sur-ajustement et améliorer le modèle et trouver une bonne architecture. La figure IV.6 présente les résultats de notre modèle CNN, pour un nombre d'époques égal à 50 et ceci avec les deux ensembles de données: apprentissage et test (train et test).

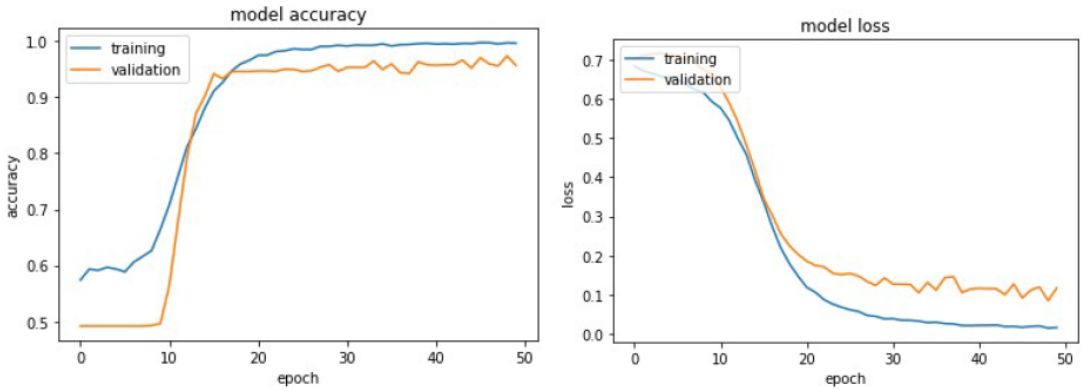


Figure IV.8 : Apprentissage et Validation modèle CNN

Comme nous pouvons le constater clairement, notre réseau CNN proposé a réussi son objectif qui est la classification des images de notre base de données, tout en gardant un taux de précision nettement élevé et une perte réduite. D'après la figure IV.8 la précision de l'apprentissage augmente avec le nombre d'époque, jusqu'à elle est stable, ceci reflète qu'à chaque époque le modèle apprenne plus d'information et la même chose pour la validation. L'erreur d'apprentissage et validation est diminuée avec le nombre d'époque. La précision de la validation est stable au tour de 94% après 20 époques par rapport aux modèles de l'apprentissage par transfert qui nécessite un grand nombre d'époques pour entraine le modèle sur notre ensemble des images jusqu'à obtenir des résultats satisfaisants avec une accuracy de 98.78.

4.2 Paramètre d'évaluation du réseau CNN

Le meilleur paramètre d'évaluation du modèle est de maximiser la précision dite « accuracy ». La fonction loss choisie pour notre modèle est binairy_crossentropy puisque nous avons choisi de faire une classification binaire. L'optimiseur qui convient à notre configuration est défini par 0.0001 pour le taux d'apprentissage (learning rate). Entre autres,

nous avons essayé d'utiliser la matrice de confusion (confusion matrix) comme paramètre d'évaluation du modèle avec la précision de test « accuracy ». comme le montre la figure IV.9

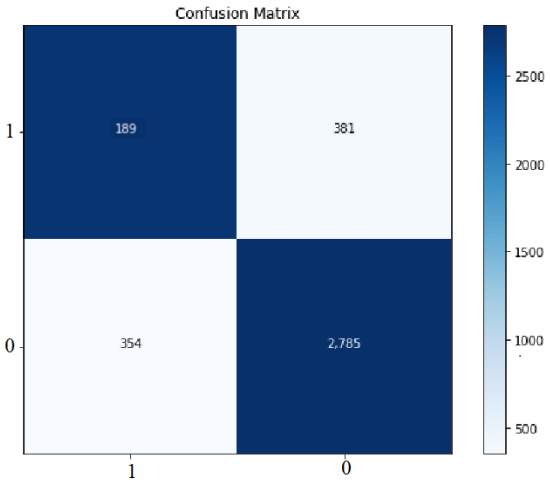


Figure IV.9 : matrice de confusion de notre CNN

4.3 Résultats obtenues de Modèle apprentissage par transfert

Nous allons créer un modèle de l'apprentissage par transfert qui sert à développer notre approche de classification des images. Tout d'abord nous allons créer une fonction pour lire la base et le redimensionner selon le besoin de l'entrée de modèle, puis nous appliquons la normalisation sur les images selon la fonction «float32 » et le diviser sur 255 pour obtenir le vecteur de 0 et 1 selon le nombre pixels de chaque image. La base sera diviser en 80% pour la l'apprentissage et 20% pour la validation selon la fonction « split ». L'avantage de ce modèle et d'avoir appliqué des modèles pré-entrainer sur notre base pour améliorer les résultats de classification en chargeant le modèle d'après keras et en spécifiant les deux paramètres. Le pois de image net pour notre modèle **weights="imagenet"** et **include_top=False (faux),** la valeur « faux » pour ne pas inclure les couches de sortie car on a deux sortie et le modèle est pré-entraine sur 1000 sortie (classe). Ce modèle surveille une quantité des données et si aucune amélioration n'est constatée pour un nombre d'époques «patience», le taux d'apprentissage est réduit. Pour exécuter le modèle sur l'ensemble de données complet, nous avons utilisé des durées différentes pour l'apprentissage et la mise à jour des poids pendant plusieurs périodes. Parce que plus le nombre d'époques est élevé, plus le modèle donne de meilleurs résultats. Après un certain temps, le modèle cesse de s'améliorer si les résultats restent les mêmes. Les figures IV.10, IV.11 et IV.12 montrent les résultats des modèles VGG16, InceptionV3, ResNet.

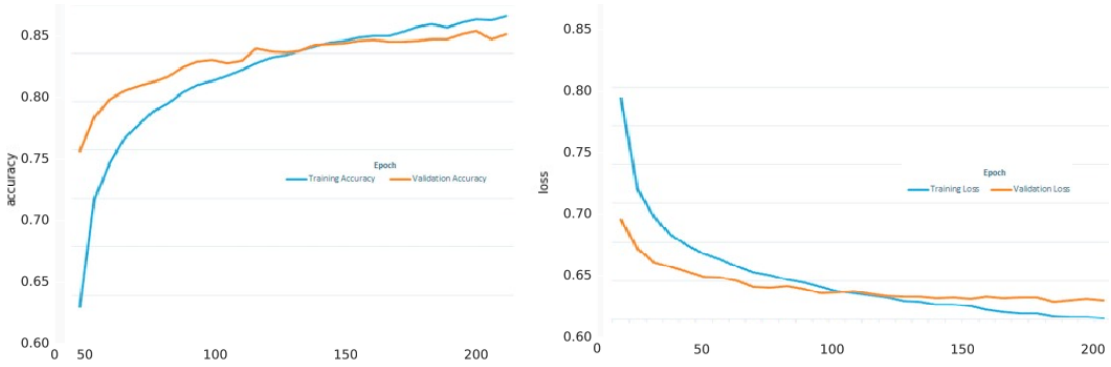


Figure IV.10 : Apprentissage et Validation modèle VGG16

D'après la Figure IV.10 (VGG16) la précision de l'apprentissage augmente avec le nombre d'époque jusqu'à elle devient stable, et ca à l'époque 150, par contre la validation augmente avec le nombre d'époque. Pour l'erreur d'apprentissage, elle diminue et se stabilise à l'époque 190 entre 0.60 et 0,63.

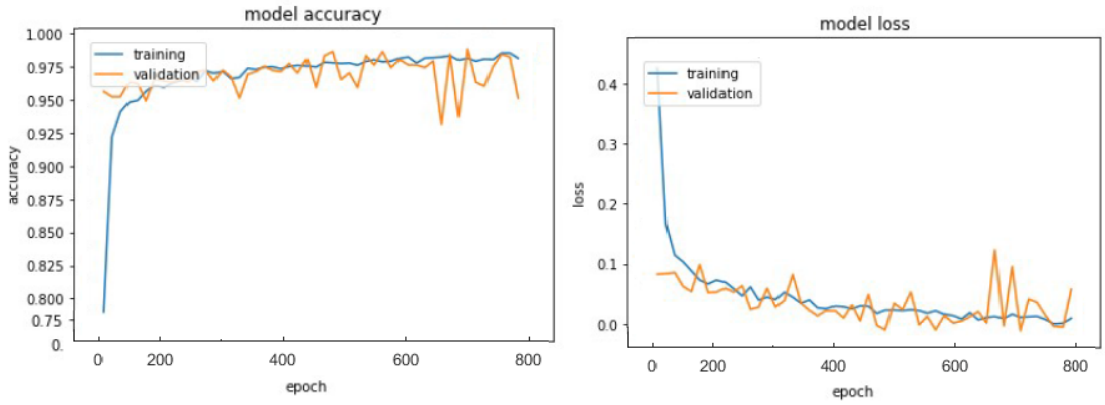


Figure IV.11 : Apprentissage et Validation modèle ResNet

D'après la Figure IV.11 (ResNet) la précision de l'apprentissage augmente avec le nombre d'époque jusqu'à elle devient stable, et ca à l'époque 200, par contre la précision de la validation est entre 0.92 et 0.95 dans tout l'exécution. L'erreur d'apprentissage est diminuée et se stabilisée a l'époque 400 entre 0 et 0,1.

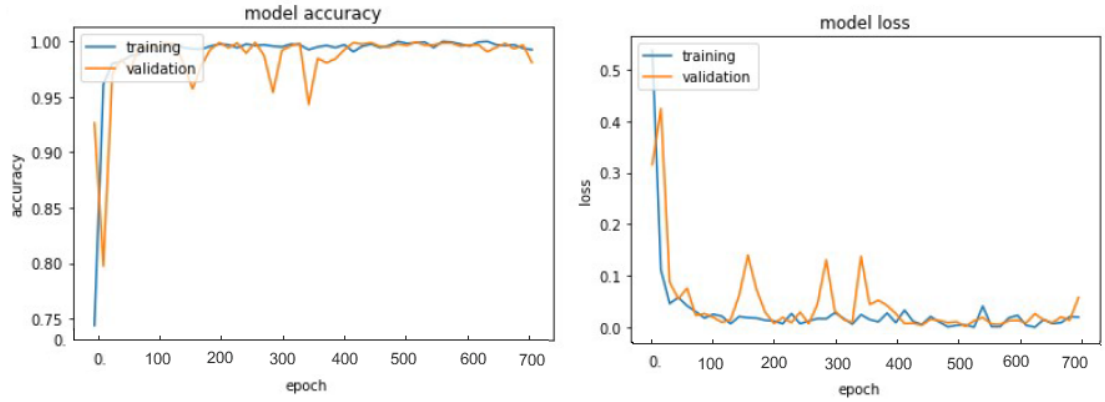


Figure IV.12 : Apprentissage et Validation modèle InceptionV3

D'après la Figure IV.12 (inceptionV3) la précision de l'apprentissage augmente avec le nombre d'époque jusqu'à sa stabilisation, par contre la validation elle est stable jusqu'à l'époque 450, elle diminue puis elle augmente et même pour l'erreur d'apprentissage est diminuée et se stabilisée à l'époque 450 entre 0 et 0,1.

Les modèles pré-entraîne pour la classification des images (Inceptionv3, VGG16 et ResNet), sont donnés des valeurs de précision encourageants, cela permet d'ouvrir de nouveaux espoirs dans le domaine de la recherche de détection des maladies, ainsi plus de profondeur et de gain de temps et d'argent.

La performance du modèle Inception V3 obtenue sur notre base est de 96 %, dont il y a pas une grande différence par rapport à la performance du modèle ResNet qui est de 94% sur la même base, bien que le nombre de paramètres traité dans ResNet soient similaires. ResNet est développé et formé dans la base de l'architecture InceptionV3. La performance du modèle VGG16 obtenue sur notre base 83 % relativement faible par rapport aux autres modèles. Mais ca reste un résultat honorable qui stimule la recherche et le développement dans ce domaine.

A travers des expériences précédentes d'étude de la contribution des coefficients cepstraux fréquentiels, et l'évaluation de l'algorithme de classification réseau de neurones convolutifs, Les meilleurs résultats sont obtenus lorsque nous mappons à la fois les paramètres cepstraux dans une image 2D et utilisons des techniques d'apprentissage en profondeur avec l'algorithme CNN sur l'ensemble d'images qui permettent l'extraction de nombreuses fonctionnalités et obtiennent des résultats satisfaisants légèrement meilleurs que ceux où ils utilisent de grandes quantités de données et une puissance de calcul.

4.4 L'évaluation

Il y a plusieurs métriques d'évaluation qui existe pour évaluer les modèles de classification des images. Nous avons pris en compte les mesures de performances suivantes :

Précision(Accuracy): Elle nous indique la précision de notre modèle. C'est le rapport entre le nombre des résultats correct (les résultats correct positifs et les résultats correct négatif) au nombre total de la population de la base des données, rapport de classification (Classification report), et matrice de confusion.

Nous présentons les résultats de l'évaluation des notre modèles d'apprentissage par transfert dans les points suivants :

- **Accuracy :** D'après le tableau IV.3, nous avons présenté les résultats de notre architecture du modèle CNN avec les autres architectures de l'apprentissage par transfert, nous avons obtenu une test-accuracy convergent avec le modèle ResNet mais avec plus de nombre de paramètres et plus de nombre de l'époque. Et le modèle inceptionV3 donne le meilleur accuracy avec 96%.

Nom_Modèle	Modèle CNN de base « Model_5 »	VGG16	InceptionV3	ResNet
Accuracy Obtenu	94 %	83 %	96 %	94 %

Tableau IV.2 : Accuracy des Modèles

- **Rapport de classification**

Nous présentons les rapports de classification des modèles (InceptionV3, ResNet, VGG16) dans le tableau IV.4. Tout d'abord, la précision est la capacité d'un classificateur à ne pas étiqueter une instance positive qui est en fait négative. Et il donne le rapport des vrais positifs à la somme d'un vrai positif et d'un faux positif. Le rappel est la capacité d'un classificateur à trouver toutes les instances positives pour chaque classe. C'est le rapport des vrais positifs à la somme des vrais positifs et des faux négatifs. Le score F1 est une moyenne harmonique pondérée de précision et de rappel tel que le meilleur score est de 1,0 et le pire est de 0,0. Macro-moyenne (Macro-avg) est calculé la métrique indépendamment pour chaque classe, puis prend la moyenne (donc toutes les classes sont traitées de manière égale), tandis qu'une micro-moyenne agrégera les contributions de toutes les classes pour calculer la métrique moyenne. Les résultats des nos modèles présenter dans ce tableau :

	Classe	Precision	Recall	F1-score
ResNet	0	1.00	0.86	0.92
	1	0.80	1.00	0.89
Macro-avg		0.83	0.93	0.93
Weighted- avg		0.76	0.90	0.93
InceptionV3	0	0.90	0.82	0.86
	1	0.89	1.00	0.99
Macro-avg		0.82	0.86	0.92
Weighted- avg		0.86	0.83	0.92
VGG16	0	1.00	0.92	0.80
	1	0.86	1.00	0.80
Macro-avg		0.82	0.82	0.82
Weighted- avg		0.82	0.82	0.83

Tableau IV.3: Rapport de classification des modèles

➢ **Matrice de confusion**

Pour les résultats de troisième performance de l'évaluation des modèles présenter dans les figures suivantes :

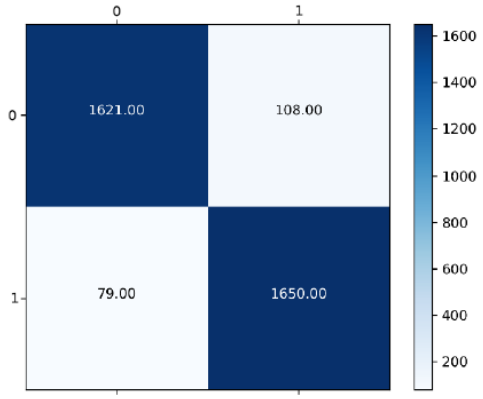


Figure IV.13 : Matrice de confusion InceptinV3

Comme le montre la matrice de confusion (Figure IV.13), pour la classe PWP (1), le modèle a pu identifier 1650 images malades et 79 images comme Non malades. On peut voir que pour la classe d'images (Non malade '0'), le modèle Inception_V3 était capable de prédire 1621 images comme Non malade et 108 comme malade.

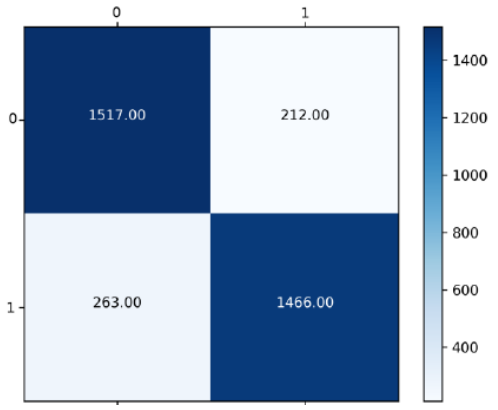


Figure IV.14 : Matrice de confusion ResNet

La matrice de confusion (Figure IV.14) indique que, pour la classe d'images (Non malade '0'), 1703 images ont été correctement prédites comme normales et 26 ont été étiquetées comme malade (1). Et pour la deuxième classe d'images (PWP), le modèle a pu identifier correctement 1638 images, mais 91 images ont été étiquetées comme non malade.

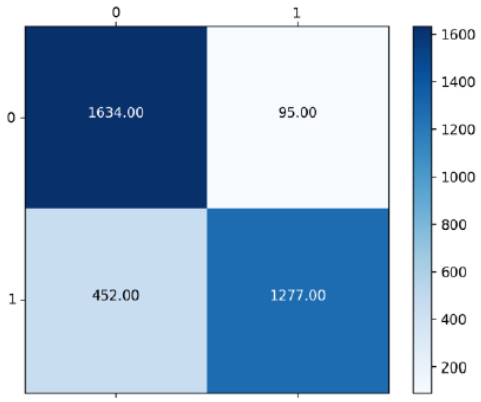


Figure IV.15: Matrice de confusion VGG16

À partir de la matrice de confusion (Figure IV.15), il est observé que la première classe d'images (Non malade '0'), le modèle a pu identifier correctement 1634 images dans la classe 0, mais 95 ont été étiquetées comme malade (classe 1). De même, pour la deuxième classe d'images (PWP '1'), le modèle a pu identifier correctement 1277 images, mais 252 images ont été étiquetées comme (non malade '0').

5. Conclusion

Dans ce dernier chapitre nous avons étudié les performances de la méthode de segmentation audio et les performances de la méthode d'extraction de caractéristiques MFCCs de la base de données des images obtenues. Ainsi, nous avons présenté d'une manière détaillée, les outils utilisés pour réaliser le travail expérimental et fournit les résultats obtenus avec une analyse et une comparaison des résultats avec les différents modèles à implémenter sur notre base

Conclusion

Générale

Conclusion Générale

Les troubles neurologiques affectant des millions d'adultes dans le monde plus précisément la maladie de Parkinson, peuvent affecter l'intelligibilité de la parole. La discrimination automatique d'une pathologie de la parole peut contribuer à l'amélioration de la communication des patients dans leur entourage. A cet effet, une étude basée sur le traitement de la voix ainsi que la classification qui pourra être utilisé comme outil pour améloirer le travail.

En premier lieu, une analyse de la parole basée sur la conversion du son en image a été présenté, afin d'augmenter la quantité de données et extraire les nombreuses fonctionnalités qui existent dans les images. Pour garantir plus de précision du système, une intégration de d'un apprentissage automatique avec plus de données a été mise en ouvre comme second lieu. Cela permettra à l'apprentissage des valeurs de référence de se poursuivre, puis de mettre à jour la base de données qui sera plus bénéfique pour améliorer la précision de la reconnaissance.

Ce dernier pourra être bénéfique pour les personnes qui souffrent d'autres divers troubles de la parole, qui ont réduit la clarté de leur voix, en particulier la maladie parkinson.

De plus, le domaine de la classification d'images a connu un grand développement depuis l'émergence du Deep-Learning. Dans ce projet que nous avons proposé une approche pour détecter l'état du patient, qu'il soit malade ou non. L'idée s'appuie sur l'utilisation des réseaux de neurones convolutifs. Principalement pour classer les images graphiques extraites, contenants des courbes qui représentent les coefficients calculés.

Afin de valider notre approche, nous avons utilisé une base de données d'images, construites de fichiers audio, et des données agrandies grâce à l'extraction des caractéristiques à travers les Mel-Frequency Cepstral Coefficients (MFCCs) et la technique de séparation des courbes représentant le ton de la voix et le chemin de la parole. Après un bon entrainement de notre modèle, on a fait les tests afin d'obtenir de meilleurs résultats en termes de précision et d'erreur. À la fin nous avons comparé notre modèle CNN avec d'autres modèles. Ces derniers sont les modèles de l'apprentissage par transfert et nous avons obtenu de bons résultats de classification autours de 96%. Ces résultats sont considéré encourageants et peuvent ouvrir des nouveaux espoirs pour développer les recherches même au cas du manque de données audio, afin de faire une propre évaluation de la santé des patients, et le développement du domaine de la médecine et sa prospérité.

En conclusion, le travail que nous avons effectué nous a permis de réaliser par des algorithmes de deep learning, Les résultats obtenus sont encourageants mais nous suggérons de continuer la recherche dans ce sens pour améliorer encore plus les résultats surtout pour arriver à une dilatation optimale qui nous permis un suivi presque parfait.

Références

Bibliographiques

Références

Akujuobi, U., & Zhang, X. (n.d.). *Delve : A Dataset-Driven Scholarly Search and Analysis System*.

Al-Fatlawi, A. H., Jabardi, M. H., & Ling, S. H. (2016). Efficient diagnosis system for Parkinson's disease using deep belief network. *2016 IEEE Congress on Evolutionary Computation, CEC 2016*, 1324–1330. https://doi.org/10.1109/CEC.2016.7743941

Arnold, L., Paugam-Moisy, H., & Sebag, M. (2010). Optimisation de la Topologie pour les Réseaux de Neurones Profonds. *17e Congrès Francophone AFRIF-AFIA Reconnaissance Des Formes et Intelligence Artificielle - RFIA 2010*.

Babii, H., & Ing, D. (1975). *until 1976*. *DECEMBER 1916*.

Bankar, J., & Gavai, N. R. (2018). Convolutional Neural Network Based Inception V3 Model for Animal Classification. *International Journal of Advanced Research in Computer and Communication Engineering*, *7*(5), 142–146. https://doi.org/10.17148/IJARCCE.2018.7529

Bar, Y., Diamant, I., Wolf, L., & Greenspan, H. (2015). Deep learning with non-medical training used for chest pathology identification. *Medical Imaging 2015: Computer-Aided Diagnosis*, *9414*, 94140V. https://doi.org/10.1117/12.2083124

Beel, J., Carevic, Z., Schaible, J., & Neusch, G. (2017). *RARD : The Related-Article Recommendation Dataset RARD : The Related-Article Recommendation Dataset*. *June*. https://doi.org/10.1045/july2017-beel

Blond, B. (2010). 5/5 La maladie de Parkinson. *Soins*, *55*(747), 43–44. https://doi.org/10.1016/s0515-3700(09)70508-7

Bock, K., & Garnsey, S. M. (2008). Advances in Language Processing. *A Companion to Cognitive Science*, 226–234. https://doi.org/10.1002/9781405164535.ch14

Buccale, C. (2015). *Cavite buccale*. 1–18.

Busto, P. P., & Gall, J. (2017). Open Set Domain Adaptation. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 754–763. https://doi.org/10.1109/ICCV.2017.88

Camps, J., Samà, A., Martín, M., Rodríguez-Martín, D., Pérez-López, C., Moreno Arostegui, J. M., Cabestany, J., Català, A., Alcaine, S., Mestre, B., Prats, A., Crespo-Maraver, M. C., Counihan, T. J., Browne, P., Quinlan, L. R., Laighin, G., Sweeney, D., Lewy, H., Vainstein, G., … Rodríguez-

Molinero, A. (2018). Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowledge-Based Systems*, *139*, 119–131. https://doi.org/10.1016/j.knosys.2017.10.017

Cao, Z., Long, M., Wang, J., & Jordan, M. I. (n.d.). *Cao_Partial_Transfer_Learning_CVPR_2018_paper.pdf*. 2724–2732.

Che, C., Xiao, C., Liang, J., Jin, B., Zhou, J., & Wang, F. (2017). An RNN architecture with dynamic temporal matching for personalized predictions of Parkinson's disease. *Proceedings of the 17th SIAM International Conference on Data Mining, SDM 2017*, 198–206. https://doi.org/10.1137/1.9781611974973.23

Cheang, H. S., & Pell, M. D. (2007). An acoustic investigation of Parkinsonian speech in linguistic and emotional contexts. *Journal of Neurolinguistics*, *20*(3), 221–241. https://doi.org/10.1016/j.jneuroling.2006.07.001

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 1800–1807. https://doi.org/10.1109/CVPR.2017.195

Cid, Y. D., Kalinovsky, A., Liauchuk, V., Kovalev, V., & Müller, H. (2017). Overview of the ImageCLEF 2017 tuberculosis task - Predicting tuberculosis type and drug resistances. *CEUR Workshop Proceedings*, *1866*.

Cisonni, J. (2008). *Modélisation et inversion d ' un système complexe de production de signaux acoustiques Application à la voix et aux pathologies*. https://tel.archives-ouvertes.fr/tel-00363714/document

Combescure, P., Le Guyader, A., Jouvet, D., & Sorin, C. (1995a). Le traitement du signal vocal. *Annales Des Telecommunications/Annals of Telecommunications*, *50*(1–2), 142–164.

Combescure, P., Le Guyader, A., Jouvet, D., & Sorin, C. (1995b). Le traitement du signal vocal voice signal processing. *Annales Des Télécommunications*, *50*(1), 142–164. https://doi.org/10.1007/BF03000774

Dai, W., Yang, Q., Xue, G. R., & Yu, Y. (2007). Boosting for transfer learning. *ACM International Conference Proceeding Series*, *227*, 193–200. https://doi.org/10.1145/1273496.1273521

Daphne, E. J. (2018). Parkinson's disease detection and classification using machine learning and deep learning algorithms-a survey. *International Journal of Engineering Science Invention (IJESI)*, *7*(5), 56–63. www.ijesi.org%7C%7CVolumewww.ijesi.org56%7C

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. Journal of speech and hearing research, 12(3), 462-496.

Davis, S. B., & Mermelstein, P. (1980). Comparison of Parametric Representations for. *Transactions on Acoustics, Speech, and Signal Processing*, *28*(4), 357–366.

Deep, D., & Tensorflow, L. (2018). *2 Installation des librairies sous R 3 Perceptron simple et multicouche*. 1–13.

Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, *3*(January 2014). https://doi.org/10.1017/atsip.2013.9

Deng, L., & Yu, D. (2013). Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, *7*(3–4), 197–387. https://doi.org/10.1561/2000000039

Dias, S. (2012). *Estimation of the glottal pulse from speech or singing voice*. 92.

Drygajlo, A. (2014). *Dr . Andrzej Drygajlo*. *January 1997*.

Frid, A., Kantor, A., Svechin, D., & Manevitz, L. M. (2017). Diagnosis of Parkinson's disease from continuous speech using deep convolutional networks without manual selection of features. *2016 IEEE International Conference on the Science of Electrical Engineering, ICSEE 2016*, *October 2017*. https://doi.org/10.1109/ICSEE.2016.7806118

Ghio, A. (2007). L'évaluation acoustique. *Les Dysarthries*, 111–114.

Godino-Llorente, J. I., Gmez-Vilda, P., & Lee, T. (2009). Analysis and signal processing of oesophageal and pathological voices. *Eurasip Journal on Advances in Signal Processing*, *2009*. https://doi.org/10.1155/2009/283504

Goetz, C. G. (2011). The history of Parkinson's disease: Early clinical descriptions and neurological therapies. *Cold Spring Harbor Perspectives in Medicine*, *1*(1), 1–15. https://doi.org/10.1101/cshperspect.a008862

Guerfi, S. (2011). *Authentification d ' individus par reconnaissance de caractéristiques biométriques liées aux visages 2D / 3D Souhila Guerfi To cite this version : HAL Id : tel-00623243*.

Hachesu, P. R., Moftian, N., Dehghani, M., & Soltani, T. S. (2017). *Analyzing a Lung Cancer Patient Dataset with the Focus on Predicting Survival Rate One Year after Thoracic Surgery*. *18*, 1531–1536. https://doi.org/10.22034/APJCP.2017.18.6.1531

Harar, P., Galaz, Z., Alonso-Hernandez, J. B., Mekyska, J., Burget, R., & Smekal, Z. (2018). Towards robust voice pathology detection. *Neural*

Computing and Applications, *July*. https://doi.org/10.1007/s00521-018-3464-7

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Henrich Bernardoni, N., & Acher, A. (2014). Comment analyser la voix humaine dans la parole et dans le chant ? Les outils scientifiques et méthodes de la recherche fondamentale à disposition de la recherche clinique sur la voix et leurs implications en orthophonie. *Rééducation Orthophonique*, *257*, 155–176.

Hopper, G. (1977). Apprendre à programmer en Python. *The Biblical Archaeologist*, *40*(2), 92. https://doi.org/10.2307/3209493

Ittichaichareon, C. (2012). Speech recognition using MFCC. *... Conference on Computer ...*, 135–138. https://doi.org/10.13140/RG.2.1.2598.3208

Katzenshlager, R., Evans, A., Manson, A., Palsalos, P. N., Ratnaraj, N., Watt, H., Timmermann, L., Van Der Giessen, R., & Lees, A. J. (2004). Mucuna pruriens in Parkinson's disease: A double blind clinical and pharmacological study. *Journal of Neurology, Neurosurgery and Psychiatry*, *75*(12), 1672–1677. https://doi.org/10.1136/jnnp.2003.028761

Kibleur, A., & George, M. N. (2016). *Cartographie corticale par électroencéphalographie des effets de la stimulation cérébrale profonde chez les patients souffrant de troubles psychiatriques réfractaires et les patients parkinsoniens* . 0–218.

Klabunde, R. (2002). Daniel Jurafsky/James H. Martin: Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. *Zeitschrift Fur Sprachwissenschaft*, *21*(1), 134–135. https://doi.org/10.1515/zfsw.2002.21.1.134

Laprie, Y. (2000). *Analyse spectrale de la parole*.

Leclercq, N. (2013). *Mémoire présenté devant le jury de l ' ENSAE en vue de l ' obtention du Diplôme d ' Actuaire ENSAE et de l ' admission à l ' Institut des Actuaires*. *33*(0).

Magdelaine-beuzelin, C. (2010). Université françois - rabelais de tours. *Management*, 476.

Massida, Z., Barone, P., & Deguine, O. (2010). Étude de la perception de la voix chez le patient sourd postlingual implanté cochléaire unilatéral et le normo-

entendant en condition de simulation d'implant. Psychophysique et imagerie. *Neurosciences*, 348.

Moerman, M., Martens, J. P., Crevier-Buchman, L., De Haan, E., Grand, S., Tessier, C., Woisard, V., & Dejonckere, P. (2006). The INFVo perceptual rating scale for substitution voicing: Development and reliability. *European Archives of Oto-Rhino-Laryngology*, *263*(5), 435–439. https://doi.org/10.1007/s00405-005-1033-z

Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T. A., Farahat, M., Malki, K. H., Al-nasheri, A., & Bencherif, M. A. (2017). Voice pathology detection using interlaced derivative pattern on glottal source excitation. *Biomedical Signal Processing and Control*, *31*. https://doi.org/10.1016/j.bspc.2016.08.002

Nitisha, & Bansal, A. (2012). Speaker Recognition Using MFCC Front End Analysis and VQ Modeling Technique for Hindi Words using MATLAB. *International Journal of Computer Applications*, *45*(24), 48–52.

Pinto, S., Ghio, A., Teston, B., & Viallet, F. (2010). Dysarthria across Parkinson's disease progression. Natural history of its components: Dysphonia, dysprosody and dysarthria. *Revue Neurologique*, *166*(10), 800–810. https://doi.org/10.1016/j.neurol.2010.07.005

Qian, Y., Dong, J., Wang, W., & Tan, T. (2018). Feature learning for steganalysis using convolutional neural networks. *Multimedia Tools and Applications*, *77*(15), 19633–19657. https://doi.org/10.1007/s11042-017-5326-1

Quellet, S. (1995). Cancer du larynx. *The Canadian Nurse*, *91*(9), 21.

Rachedi, J., Sciences, M., Sp, U., Parcours, S. A. R., Laboratoire, A., Equipe, P., Temps, A., Schnell, N., & Mars, D. S. (2005). *Reconnaissance et classification de phon ` emes*.

Rhonda, J. H., Oates, J. M., Debbie, J. P., & Andrew, J. H. (2000). Voice characteristics in the progression of Parkinson's disease. *International Journal of Language and Communication Disorders*, *35*(3), 407–418. https://doi.org/10.1080/136828200410654

Ritchings, R. T., McGillion, M., & Moore, C. J. (2002). Pathological voice quality assessment using artificial neural networks. *Medical Engineering and Physics*, *24*(7–8), 561–564. https://doi.org/10.1016/S1350-4533(02)00064-4

Rosso, C. De, Du, E. T., Master, G. D. E., & Caroline, D. R. (2018). *Évaluation acoustique de la voix de patients présentant des symptômes d ' anxiété*

généralisée To cite this version : HAL Id : dumas-01895594 DE PATIENTS PRESENTANT DES SYMPTOMES D ' ANXIETE GENERALISEE.

S. Young, G. E., T. Hain, D. Kershaw, X. Liu, G. Moore, J.Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland. (1997). The HTK Book (for HTK Version 3.4). Cambridge University Engineering Department.

Shahi, T. B., & Pant, A. K. (2018). Nepali news classification using Naïve Bayes, Support Vector Machines and Neural Networks. *Proceedings - 2018 International Conference on Communication, Information and Computing Technology, ICCICT 2018*, *2018-Janua*, 1–5. https://doi.org/10.1109/ICCICT.2018.8325883

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–14.

Sun, J., Chong, P., Tan, Y. X. M., & Binder, A. (2017). ImageCLEF 2017: ImageCLEF tuberculosis task - The SGEast submission. *CEUR Workshop Proceedings*, *1866*.

Sun, R., Limkin, E., Dercle, L., Reuzé, S., Chargari, C., Schernberg, A., Dirand, A., Alexis, A., Paragios, N., Sun, R., Limkin, E., Dercle, L., Reuzé, S., & Zacharaki, E. (2017). *Imagerie médicale computationnelle (radiomique) et potentiel en immuno-oncologie To cite this version : HAL Id : hal-01668902*.

Sundberg, J. (1954). Acoustics of the singing voice. *Nature*, *174*(4433), 716–717. https://doi.org/10.1038/174716a0

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 2818–2826. https://doi.org/10.1109/CVPR.2016.308

Tan, P. (2017). *Précision de modèle et efficacité algorithmique : exemples du traitement de l ' occultation en stéréovision binoculaire et de l ' accélération de deux algorithmes en optimisation convexe To cite this version : HAL Id : tel-01420603 par Pr ′ ecision de mod*.

Tarayoun, T. (2018). Méthodes de classification supervisée appliquées aux

terres agricoles - Résumé. *Journées de Méthodologie Statistique*. http://jms-insee.fr/jms2018_programme/%0Ahttp://jms-insee.fr/jms2018s08_3/

Teston, B. (2007). *L ' étude instrumentale des gestes dans la production de la parole : Importance de l ' aérophonométrie . To cite this version : HAL Id : hal-00173547*.

TUBACH, J., CHOLLET, G., CHOUKRI, K., MONTACIE, C., MOKBEL, C., & VALBRET, H. (1990). Adaptation au locuteur de systèmes de reconnaissances. Régression linéaire multiple et perceptrons multicouches. *TS. Traitement Du Signal*, *7*(4), 285–292.

Une, D. V. G. (2014). *Le Deep Learning pas à pas PARTIE I T Concepts*. 1–9.

Viallet, F., Teston, B., Viallet, F., Teston, B., Parkinson, D., & Les, P. A. (2008). *La dysarthrie dans la maladie de Parkinson To cite this version : HAL Id : hal-00242418*. 169–174.

Woo, N. H. (1969). *Prosody and phonology (Unpublished doctoral dissertation)*.

Works, P., & Recognition, S. (1959). *MFCC Features*. 7–16. https://doi.org/10.1007/978-3-319-49220-9

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, *95*(4), 2240–2253. https://doi.org/10.1121/1.408684

Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., & Woodland, P. (2002). *The HTK Book (for HTK Version 3.2)*. *March 1997*.

Zantis, F. (2002). *Reconnaissance vocale interactive*. 28–33.

Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. *Journal of Computer Science and Technology*, *16*(6), 582–589. https://doi.org/10.1007/BF02943243

Zheng, J., Xu, C., Zhang, Z., & Li, X. (2017). Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network. *2017 51st Annual Conference on Information Sciences and Systems, CISS 2017*, *March*. https://doi.org/10.1109/CISS.2017.7926112